

Copyright  
by  
Emery Charles Wolf  
2017

**The Thesis Committee for Emery Charles Wolf  
Certifies that this is the approved version of the following thesis:**

**POLICY NETWORK AND CONTENT ANALYSIS: APPLICATIONS  
IN WATER RESOURCES MANAGEMENT AND SCIENCE**

**APPROVED BY  
SUPERVISING COMMITTEE:**

**Supervisor:**

---

Suzanne A. Pierce

---

Richard Chuchla

---

Bart Bohn

**POLICY NETWORK AND CONTENT ANALYSIS: APPLICATIONS  
IN WATER RESOURCES MANAGEMENT AND SCIENCE**

**by**

**Emery Charles Wolf, B.S.**

**Thesis**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Arts**

**The University of Texas at Austin**

**May 2017**

## **Dedication**

This thesis is dedicated to my friends and family at Pearl Street who have simultaneously been able to provide emotional support and strife through the arduous process of working through my degree.



## **Acknowledgements**

I would primarily like to acknowledge Mr. Chuchla for being available with guidance for me during the eleventh hour. This work would not have been possible without the help of the Energy & Earth Resources Program at the University of Texas at Austin and the support of the others in my cohort. The Texas Water Development Board is also integral to this research by providing the documents needed for my dataset which are available online and publically accessible. I would also like to acknowledge my parents Mary & Karl Wolf for keeping me fed and housed through the past years of college.

## **Abstract**

# **POLICY NETWORK AND CONTENT ANALYSIS: APPLICATIONS IN WATER RESOURCES MANAGEMENT AND SCIENCE**

Emery Charles Wolf, M.A.

The University of Texas at Austin, 2017

Supervisor: Suzanne A. Pierce

**Abstract:** This study extends previous work using the state water plans from 1961-2017 with the most recent 2016 regional plan submissions from the Texas Water Development Board, to implement and evaluate a topic analysis methodology. The approach uses statistical analysis of the collection of text documents or *corpus* to evaluate. Topic Modeling is a systematic approach for analyzing the relationships, usage frequency of words and communities of words to extract themes, concepts, and informational meaning from a selected corpus.

This research documents methods for content analysis that can be used on state water plans, as well as other environmental science and policy documents. For this study, nearly 19,658 pages of text from the state and regional water plans for Texas were analyzed. Unsurprisingly, results indicate that “water” is the central common theme connecting all topics. Early results identified a set of primary topics that are shared throughout all regions including planning, strategy, and groundwater. Interestingly, themes varied from west to east reflecting the gradient of arid to humid climates respectively. In the West, themes indicate that regional water planning groups focus more heavily on irrigation and wells for agriculture, while in the East the focus tends to be for

municipal uses and surface water strategies, such as reservoirs and infrastructure. This thematic pattern also aligns with the population distribution of Texas, with larger numbers of people in the east, and much less dense populations in the west. Analyses of the state water plans over time illustrate that topics related to drought, planning, and water needs have increased over the period under study. Network statistics reveal that the largest change between state water plans occurred between the 1961 and 1968 plans. Topic analysis methodologies provide an accessible and systematic approach to evaluate the context of water planning, management, and policy across the state. The approach may provide a mechanism for linking quantitative science knowledge about water resources in the state with the qualitative planning and policy perspectives used to manage these critical resources.

## Table of Contents

List of Tables .....	ix
List of Figures .....	x
Chapter I – Introduction.....	1
Introduction of the TWDB .....	1
Study Workflows .....	2
Chapter II - Water Resources, Governance and Management.....	3
Groundwater Management Areas .....	3
Groundwater Conservation Districts.....	5
Convergence of Groundwater Management .....	7
Regional Water Planning Groups .....	9
Chapter III – Background Literature for Content and Network Analysis .....	11
Reproducible Frameworks .....	11
History of Quantitative Content Analysis .....	11
Previous Efforts Using Informatics on State Water Plans .....	14
Chapter IV – Methodology .....	15
Preprocessing and Preparing the Dataset for Analysis .....	15
Word Counts and Frequency Analyses .....	16
Topic Modeling in Rstudio .....	18
Indexed Progression of Common Topics .....	20
Commercial Concept Mapping Tool: Leximancer .....	24
Topic modeling in networks with Texttexture and Gephi .....	29
Statistical Analysis of State Water Plans Topic Networks .....	34
Regional Water Plans Represented by the State Plan .....	35
Chapter V – Discussion .....	37
State Water Plans .....	37
Regional Water Plans.....	40
Reproducible Framework.....	40
Chapter VI – Conclusions.....	43
APPENDICES .....	46
APPENDIX A - FREQUENCY ANALYSIS: HISTOGRAMS AND WORDCLOUDS OF STATE WATER PLANS 1961-2017 .....	46
APPENDIX B - CLUSTER ANALYSIS OF REGIONAL WATER PLANS PROJECTED THROUGH LEXIMANCER .....	52
APPENDIX C - TWDB STATE WATER PLAN NETWORK VISUALIZATIONS USING GEPHI.....	56
APPENDIX D - TEXAS GROUNDWATER MANAGEMENT AREAS FUTURE DESIRED CONDITIONS TABLE .....	66
APPENDIX E – R CODE FOR CLEANING TEXT FILES AND FREQUENCY ANALYSIS AND FINDING CORRELATIONS.....	75
APPENDIX F – Leximancer Settings for Network Output .....	81
References.....	82

## **List of Tables**

Table 1. TWDB State Water Plans Statistical Analysis of Generated Networks.....	35
---	----

## List of Figures

Figure 1. Groundwater Management Areas (TWDB, 2017).....	4
Figure 2. Major Aquifers of Texas (TWDB, 2017).....	5
Figure 3. Map of Groundwater Conservation Districts of Texas (TWDB, 2017).....	6
Figure 4. Combined map of Major Aquifers Superimposed on Groundwater Management Areas of Texas (TWDB, 2017).....	9
Figure 5. Frequency Histogram and WordCloud for the Cumulative State Water Plans 1961 - 2017 .....	17
Figure 6. Words Correlated with Water in the Cumulative State Water Plans .....	20
Figure 7. Words Correlated with Drought in the Cumulative State Water Plans.....	20
Figure 8. Words Correlated with Pollution in the Cumulative State Water Plans .....	20
Figure 9. Change in Pervasiveness of Common Themes Across all State Water Plans.....	21
Figure 10. Annual Rainfall vs. Pervasiveness of Drought in the State Water Plans .....	22
Figure 11. State Population vs. Drought in the State Water Plans .....	24
Figure 12. Leximancer Network Visualization of Region E (West Texas) Regional Water Plan .....	25
Figure 13. Leximancer Network Visualization of Region D (East Texas) Regional Water Plan .....	26
Figure 14. 1961 State Water Plan Network Visualization - Visualized with Gephi .....	30
Figure 15. 2002 State Water Plan Visualization – Visualized with Gephi.....	31
Figure 16. Total Words in Each State Water Plan.....	39
Figure 17. The Policy Analysis Framework for Combining Qualitative and Quantitative Approaches to Analyze Water Policy Documents.....	41

## **CHAPTER I – INTRODUCTION**

### **Introduction of the TWDB**

The drought of the 1950's in Texas created a statewide response which was responsible for the development of the Texas Water Development Board (TWDB, 2017). The continuing goal of the TWDB as an organization is to develop water supplies while preparing to meet future water needs of Texas. Composed of 16 planning groups, the TWDB periodically publishes regional and statewide planning documents. To date, there are 10 State Water Plans, all of which are included in this research. In the intervening decades since the formation of the TWDB, there have been thousands upon thousands of pages of text created as part of the planning process, aiming to document and inform actions that will maintain water resources across the state. Understanding themes and developing topics from the large set of water planning documents is imperative for both the citizens of Texas and the Texas Legislation.

Water choices affect everyone in the state, yet just 179 people within the state government make policy choices to serve 27 million people in Texas. Few politicians are trained in hydrological sciences or engineering fields, therefore methods that help distill key information, big ideas, and critical concerns identified by stakeholders, scientists, and water managers across the state can bring value to policy making processes.

Texas has a rich history of water use from both agricultural and municipal perspectives. With the historical rule of capture for groundwater in Texas, landowners have been able to withdraw water with little limitation or oversight as long as the water is used productively and without malicious intent (TCEQ, 2014). As the demand for water and population increase, the value of formulating water policy and planning that is based on sound scientific information also increases. Well posed analyses that use transparent or reproducible approaches to understand possible strategies or choices improves the

resilience of water management for everyone in the state. The aquifer yield continuum framework provides a formulation to inform governance of groundwater withdrawals and describing sustainable yield relationships based on operational and consensus yields (Pierce et al., 2013). While groundwater availability models and other monitored or observation network data can be used to understand the operational and quantitative responses of aquifer systems, the stakeholder perspectives and regional preferences are more difficult to capture in any reproducible manner. The best sources of documented information about stakeholder preferences may be contained in state and regional water plans compiled and maintained by the Texas Water Development Board. These documents serve to provide access to important aspects of the consensus yield side of this equation. With the historical records combined with governance processes leveraging planning and science-based models, Texas has a rich dataset of information to analyze. The recurring state water plans delineate primary topics or concerns as they change over time while the regional water plans give insight into how priority issues are distinct between different communities and geospatially.

## **Study Workflows**

This study uses semantic and content analysis workflows to analyze contextual traces and patterns documented in the water resource management texts of Texas. The aim is to create simple but statistically significant representations of these plans, so that the core topics and patterns can be easily understood by a wide range of audiences, as well as to understand how water management changes over time and space.

With a sufficient database of water plans in Texas, as well as plenty of literature on text mining methods, there has yet to be much research done with text mining of the water plans of Texas, which makes this research both novel and useful for further research investigations.



## CHAPTER II - WATER RESOURCES, GOVERNANCE AND MANAGEMENT

### Groundwater Management Areas

Groundwater Management Areas (GMA's) are geographical subdivisions of the State of Texas, designed to be representative of groundwater resources which are typically do not follow traditional political boundaries (fig. 2). These GMA's which were delineated using geographical information systems by the Texas Water Development Board with an attempt to follow hydrogeological boundaries. The figure below shows the State of Texas with the 16 GMAs outlined, which cross political boundaries of county lines. Figure 2 below this shows the major aquifers of Texas.

Sixteen Groundwater Management Areas are tasked "*...to provide for the conservation, preservation, protection, recharging, and prevention of waste of the groundwater, and of groundwater reservoirs or their subdivisions, and to control subsidence caused by withdrawal of water from those groundwater reservoirs or their subdivisions*" ([Texas Water Code §35.001](#)). Figure 4 presents a map that overlays the major aquifers of Texas with the delineations of the Groundwater Management Areas.

GMA boundaries generally align with major aquifer boundaries and unlike GCD's, GMA's encompass the entirety of the state. The largest of the Texas aquifers (Gulf Coast, Carrizo-Wilcox, and Ogallala) are divided into multiple GMA's due to the vast geographical areas of land that they take up, creating the potential for different needs in different areas of the same aquifers.

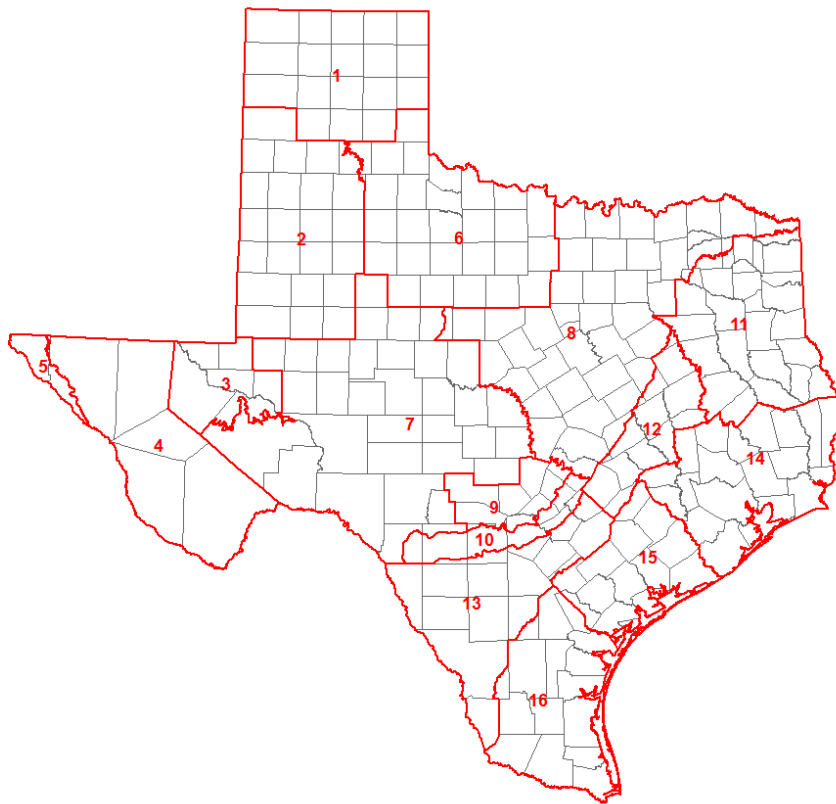


Figure 1. Groundwater Management Areas (TWDB, 2017).

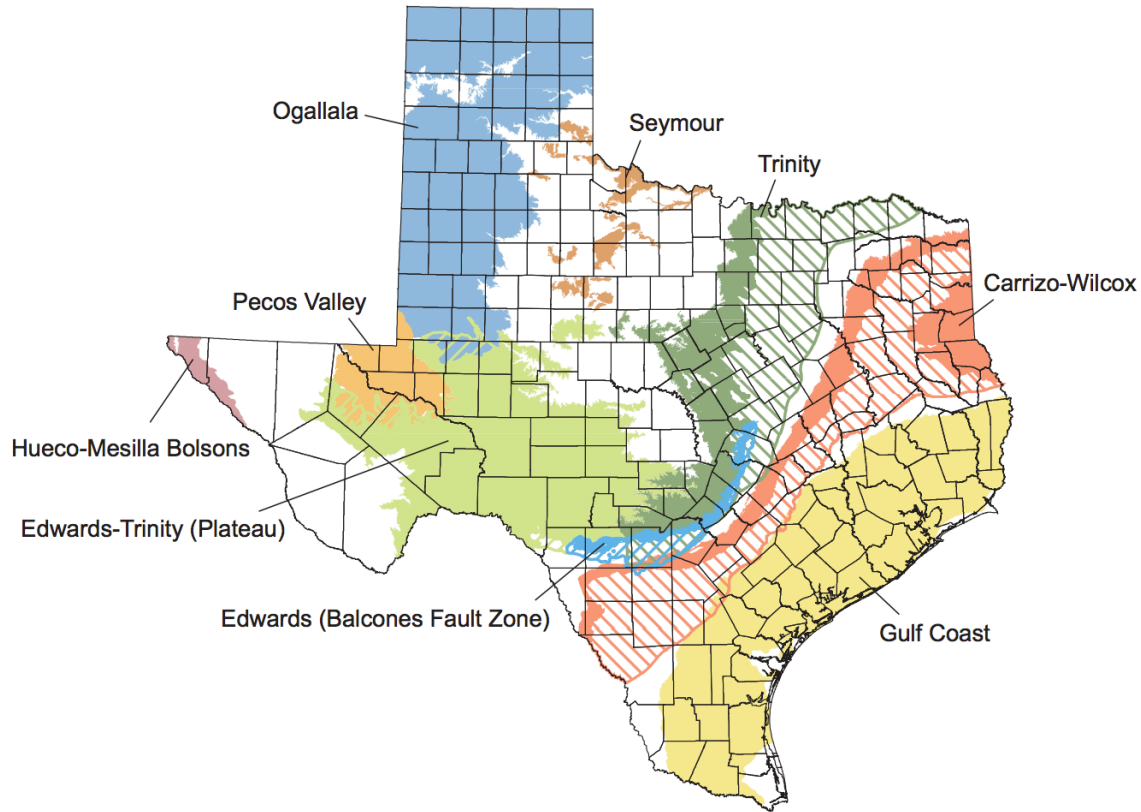


Figure 2. Major Aquifers of Texas (TWDB, 2017).

## Groundwater Conservation Districts

Texas' water is regulated by multiple agencies. Groundwater Conservation Districts (GCDs) (fig. 3) are the smallest governing entities. GCDs began in 1951, with the formation of The High Plains Underground Water Conservation District, making GCDs the oldest groundwater governance entities in the state. GCD's are one level below GMA's in the governance hierarchy. Over the next few decades, the state legislature passed additional laws aimed to encourage creation of additional districts. GCDs can be formed via four different routes; 1) action of State Legislature; usually introduced by local senator or representatives, 2) petition of property owners, where the majority of landowners in the proposed area must all sign a petition agreeing to form a conservation

**Confirmed Groundwater Conservation Districts**

1. Bender County River Authority & Ground Water District - 11/10/89  
 2. Horton Sargent Water Supply Authority - 11/10/1987  
 3. Bexar GCD - 1/20/201  
 4. Blanco-Potterville GCD - 1/23/2001  
 5. Institutional GCD  
 6. Brazoria County GCD - 11/10/2006  
 7. Brown County GCD - 11/10/2006  
 8. Brewster County GCD - 11/10/201  
 9. Brown County GCD - 11/10/2006  
 10. Callahan County GCD - 11/10/2014  
 11. Central Texas GCD - 1/23/2001  
 12. Childress County GCD - 11/10/2006  
 13. Clearwater WUCD - 8/21/2001  
 14. Coastal Bend GCD - 11/10/201  
 15. Coastal Plains GCD - 11/10/201  
 16. Coke County GCD - 11/10/1989  
 17. Colorado County GCD - 11/10/2007  
 18. Comal County GCD - 11/10/201  
 19. Comal County GCD - 11/10/201  
 20. Comal County GCD - 11/10/201  
 21. Comal County GCD - 11/10/201  
 22. Comal County GCD - 11/10/201  
 23. Comal County GCD - 11/10/201  
 24. Comal County GCD - 11/10/201  
 25. Comal County GCD - 11/10/201  
 26. Comal County GCD - 11/10/201  
 27. Comal County GCD - 11/10/201  
 28. Comal County GCD - 11/10/201  
 29. Comal County GCD - 11/10/201  
 30. Comal County GCD - 11/10/201  
 31. Comal County GCD - 11/10/201  
 32. Comal County GCD - 11/10/201  
 33. Comal County GCD - 11/10/201  
 34. Comal County GCD - 11/10/201  
 35. Comal County GCD - 11/10/201  
 36. Comal County GCD - 11/10/201  
 37. Comal County GCD - 11/10/201  
 38. Comal County GCD - 11/10/201  
 39. Comal County GCD - 11/10/201  
 40. Comal County GCD - 11/10/201  
 41. Comal County GCD - 11/10/201  
 42. Comal County GCD - 11/10/201  
 43. Comal County GCD - 11/10/201  
 44. Comal County GCD - 11/10/201  
 45. Comal County GCD - 11/10/201  
 46. Comal County GCD - 11/10/201  
 47. Comal County GCD - 11/10/201  
 48. Comal County GCD - 11/10/201  
 49. Comal County GCD - 11/10/201  
 50. Comal County GCD - 11/10/201  
 51. Comal County GCD - 11/10/201  
 52. Comal County GCD - 11/10/201  
 53. Comal County GCD - 11/10/201  
 54. Comal County GCD - 11/10/201  
 55. Comal County GCD - 11/10/201  
 56. Comal County GCD - 11/10/201  
 57. Comal County GCD - 11/10/201  
 58. Comal County GCD - 11/10/201  
 59. Comal County GCD - 11/10/201  
 60. Comal County GCD - 11/10/201  
 61. Comal County GCD - 11/10/201  
 62. Comal County GCD - 11/10/201  
 63. Comal County GCD - 11/10/201  
 64. Comal County GCD - 11/10/201  
 65. Comal County GCD - 11/10/201  
 66. Comal County GCD - 11/10/201  
 67. Comal County GCD - 11/10/201  
 68. Comal County GCD - 11/10/201  
 69. Comal County GCD - 11/10/201  
 70. Comal County GCD - 11/10/201  
 71. Comal County GCD - 11/10/201  
 72. Comal County GCD - 11/10/201  
 73. Comal County GCD - 11/10/201  
 74. Comal County GCD - 11/10/201  
 75. Comal County GCD - 11/10/201  
 76. Comal County GCD - 11/10/201  
 77. Comal County GCD - 11/10/201  
 78. Comal County GCD - 11/10/201  
 79. Comal County GCD - 11/10/201  
 80. Comal County GCD - 11/10/201  
 81. Comal County GCD - 11/10/201  
 82. Comal County GCD - 11/10/201  
 83. Comal County GCD - 11/10/201  
 84. Comal County GCD - 11/10/201  
 85. Comal County GCD - 11/10/201  
 86. Comal County GCD - 11/10/201  
 87. Comal County GCD - 11/10/201  
 88. Comal County GCD - 11/10/201  
 89. Comal County GCD - 11/10/201  
 90. Comal County GCD - 11/10/201  
 91. Comal County GCD - 11/10/201  
 92. Comal County GCD - 11/10/201  
 93. Comal County GCD - 11/10/201  
 94. Comal County GCD - 11/10/201  
 95. Comal County GCD - 11/10/201  
 96. Comal County GCD - 11/10/201  
 97. Comal County GCD - 11/10/201  
 98. Comal County GCD - 11/10/201  
 99. Comal County GCD - 11/10/201  
 100. Comal County GCD - 11/10/201

**Unconfirmed Groundwater Conservation Districts**

1. Kinney County GCD - 11/10/2001  
 2. Kinney County GCD - 11/10/2001  
 3. Live Oak GCD - 11/10/1989  
 4. Live Oak GCD - 11/10/1989  
 5. Live Oak GCD - 11/10/1989  
 6. Live Oak GCD - 11/10/1989  
 7. Live Oak GCD - 11/10/1989  
 8. Live Oak GCD - 11/10/1989  
 9. Live Oak GCD - 11/10/1989  
 10. Live Oak GCD - 11/10/1989  
 11. Live Oak GCD - 11/10/1989  
 12. Live Oak GCD - 11/10/1989  
 13. Live Oak GCD - 11/10/1989  
 14. Live Oak GCD - 11/10/1989  
 15. Live Oak GCD - 11/10/1989  
 16. Live Oak GCD - 11/10/1989  
 17. Live Oak GCD - 11/10/1989  
 18. Live Oak GCD - 11/10/1989  
 19. Live Oak GCD - 11/10/1989  
 20. Live Oak GCD - 11/10/1989  
 21. Live Oak GCD - 11/10/1989  
 22. Live Oak GCD - 11/10/1989  
 23. Live Oak GCD - 11/10/1989  
 24. Live Oak GCD - 11/10/1989  
 25. Live Oak GCD - 11/10/1989  
 26. Live Oak GCD - 11/10/1989  
 27. Live Oak GCD - 11/10/1989  
 28. Live Oak GCD - 11/10/1989  
 29. Live Oak GCD - 11/10/1989  
 30. Live Oak GCD - 11/10/1989  
 31. Live Oak GCD - 11/10/1989  
 32. Live Oak GCD - 11/10/1989  
 33. Live Oak GCD - 11/10/1989  
 34. Live Oak GCD - 11/10/1989  
 35. Live Oak GCD - 11/10/1989  
 36. Live Oak GCD - 11/10/1989  
 37. Live Oak GCD - 11/10/1989  
 38. Live Oak GCD - 11/10/1989  
 39. Live Oak GCD - 11/10/1989  
 40. Live Oak GCD - 11/10/1989  
 41. Live Oak GCD - 11/10/1989  
 42. Live Oak GCD - 11/10/1989  
 43. Live Oak GCD - 11/10/1989  
 44. Live Oak GCD - 11/10/1989  
 45. Live Oak GCD - 11/10/1989  
 46. Live Oak GCD - 11/10/1989  
 47. Live Oak GCD - 11/10/1989  
 48. Live Oak GCD - 11/10/1989  
 49. Live Oak GCD - 11/10/1989  
 50. Live Oak GCD - 11/10/1989  
 51. Live Oak GCD - 11/10/1989  
 52. Live Oak GCD - 11/10/1989  
 53. Live Oak GCD - 11/10/1989  
 54. Live Oak GCD - 11/10/1989  
 55. Live Oak GCD - 11/10/1989  
 56. Live Oak GCD - 11/10/1989  
 57. Live Oak GCD - 11/10/1989  
 58. Live Oak GCD - 11/10/1989  
 59. Live Oak GCD - 11/10/1989  
 60. Live Oak GCD - 11/10/1989  
 61. Live Oak GCD - 11/10/1989  
 62. Live Oak GCD - 11/10/1989  
 63. Live Oak GCD - 11/10/1989  
 64. Live Oak GCD - 11/10/1989  
 65. Live Oak GCD - 11/10/1989  
 66. Live Oak GCD - 11/10/1989  
 67. Live Oak GCD - 11/10/1989  
 68. Live Oak GCD - 11/10/1989  
 69. Live Oak GCD - 11/10/1989

6

## **Convergence of Groundwater Management**

GCD's within GMA's established initial desired future conditions (DFCs) for aquifers in their jurisdictions in 2010. DFCs are identified by GCDs for a 50 year planning horizon. For example, the first round of DFCs was defined for the time frame between 2010-2060. A review and summary of Desired Future Conditions for regional groups across the state are included in Appendix D. Several commonalities can be observed from comparing the DFCs in Table D-1, such as acceptable level of aquifer decline or minimum spring flow regulations. As pointed out in Robert Mace's article, GMA's were granted more power in 2005 when House Bill 1763 granted them the ability to make their own decisions on groundwater availability, require Regional Water Planning Groups (RWPG's) to use the groundwater availability numbers created by the GMA's, and finally require a permitting target of production of groundwater (Mace et al., 2008). These new powers of GMA's overflow into the role of regional planning groups, and eventually play out in the state water plans making the Desired Future Conditions (Appendix D) integral components to the creation of the state water plans.

The metrics and definitions each District uses to define and measure aquifer conditions is different, and sometimes, incompatible. A comparative look at several GMAs reflects the different priorities and approaches used to establish DFC metrics across the state. For example, the two GMAs that encompass the Ogallala Aquifer, GMA 1 and 2, are anticipating drawdown of the aquifer. Their DFCs are based upon different metrics, with some GCD's setting limits to maintain 50% of 2010 saturated thickness volumes by 2060, while others desire to not have a drawdown exceeding 50 feet in the same time period. All constituents in these districts anticipate further drawdown of the Ogallala aquifer, which has already experienced significant withdrawal but how to limit or permit change through time remains ambiguous and distinct from district to district. In comparison, other GMAs focused principally on spring flow as a performance measure for DFC definitions. GMA 7 which overlays the Edwards Aquifer, anticipates net withdrawals with drawdown estimates typically ranging between 5 to 30 feet. According to the Texas State Historical Handbook, the Los Moras Springs are the 9th largest in

Texas, and have been providing recreation for people for at least 11,000 years (Waters et al., 2008). This contrast for the use of groundwater for recreation compares with other GMAs focused on agriculture. At the southern section of the Carrizo-Wilcox Aquifer, GMA 13 also uses spring flow as a metric. The Edwards (BFZ) GCD desires at least 500 gallons per minute from producing springs in Frio County. These springs contribute to the Frio River, which is a major recreational attraction to the area. DFC design in GMA 14 reflects communication among GCD's as evidenced by closely aligned metrics of the different districts. Overlying the Gulf Coast Aquifer, this GMA aims to maintain a saturated thickness of 90%. Although these goals seem lofty in comparison with GMA 1 and 2, the Gulf Coast Aquifer has the geographical luxury of being in East Texas where precipitation is much higher, creating an inversely proportional relationship with irrigation demands as compared with many other regions in the state. GMA 10, within the Edwards Balcones Fault Zone (BFZ) is the only management area in the state which is not allowing any net drawdown of groundwater levels over the next 50 years.

In summary, DFC's vary spatially, and various stakeholders rely on different uses for the aquifers, which makes developing a comprehensive water plan challenging for officials.

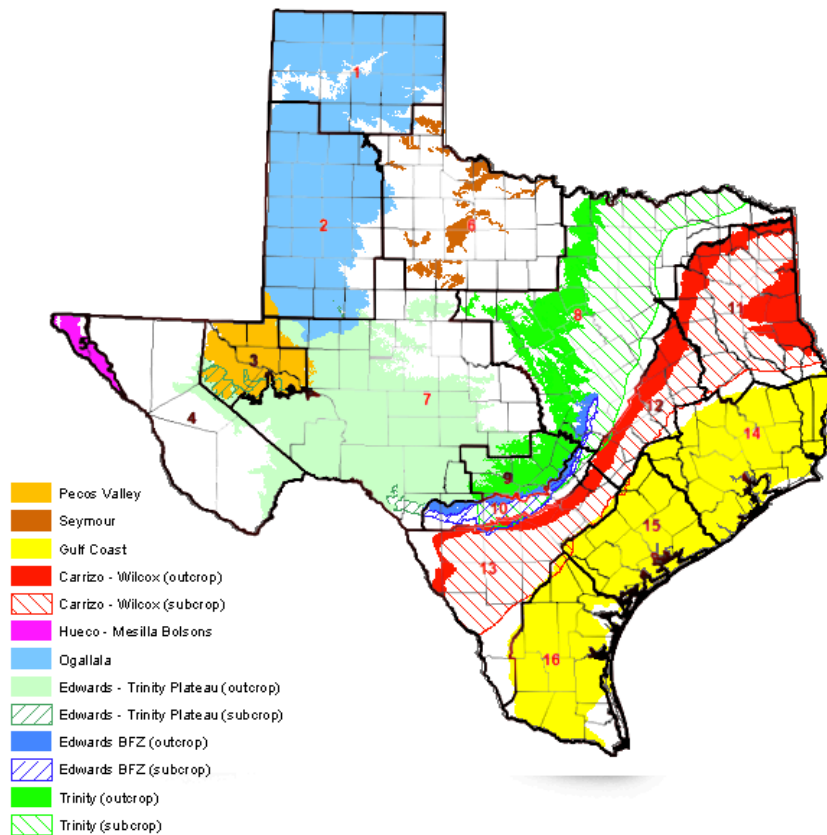


Figure 4. Combined map of Major Aquifers Superimposed on Groundwater Management Areas of Texas (TWDB, 2017).

## Regional Water Planning Groups

At the regional scale, GMAs are part of Regional Water Planning Groups (RWPGs) in Texas. These 16 regions, which are based upon political boundaries of counties, crisscross the GMA's (shown in fig. 5) and the RWPGs are responsible for surface water as well as groundwater. According to the Texas Water Development Board, the massive drought which occurred in the 1950's prompted implementation of a cyclical planning process that includes the development and continual updates and publication of state water plans.

RWPGs develop and agree upon a plan for their region, which is then sent to the TWDB for incorporation into a holistic state plan. Regional plans are aimed to satisfy a

variety of stakeholder interests, including: agriculture, industry, environment, public, municipalities, business, water districts, river authorities, water utilities, counties, groundwater management areas, and power generation (Rogers & Clancy, 2014).



## **CHAPTER III – BACKGROUND LITERATURE FOR CONTENT AND NETWORK ANALYSIS**

### **Reproducible Frameworks**

One approach for creating a reproducible framework for analyzing the water plans is to use automated content and network analysis (Wehner, 2011). These statistical methods known as content analysis, semantic analysis, or topic modeling all work by determining how reoccurring words are arranged and how they are connected to other words within a corpus or collection of textual material. Text mining methods such as these are useful in creating network visualizations of the key ideas within documents.

### **History of Quantitative Content Analysis**

Trying to use the language of math to understand natural language has gained traction with the advent of computers, but has existed for much longer. The first known attempt at using a quantitative approach for analyzing bodies of text was in 1743 when the Lutheran State Church worried that a new set of hymns *Songs of Zion*, which was gaining popularity, may have been introduced by German influences and radical new ways of thinking. Clergy set upon a mission to empirically determine if the messages within the hymns deviated from the teachings of the Lutheran Church. By hand counting mentions of religious figures within the hymns, clergy members determined that the mentions of Jesus were disproportionately larger than other mentions of God, and the Holy Ghost. This finding shined a light on the hymns, signifying that they indeed deviated from the Orthodox Lutheranism of the time (Dovring, 1954). This is accepted by many as the first documented approach of a pseudo Content Analysis.

Creating visualizations to aid in the understanding of a corpus has existed for decades. One of the first published examples was a breakdown of the classic Grimm Brothers fairy tale *Hansel & Gretel*, to uncover hidden structures which aren't as apparent by only reading through text (Bruce & Newman, 1978). This early

representation was done without the aid of the computational processing power that is available today.

Later research shifted to the use of computers, resulting in the creation of a program called BORIS that was used to analyze the effect or tone of a corpus, by identifying positive and negative emotions of characters within a corpus (Dyer, 1983). Using 400 words of emotion which were identified by hand to be positive or negative, BORIS was able to scan text and determine effect for a narrative (Dyer, 1983). Although quite limited in effectiveness, this work was some of the earliest computer aided analyses to understand forces within narrative structures. As computers continued to aid the analyses of text documents, the ability to connect concepts arose. Map Analysis together with Content Analysis was utilized to connect the concepts in a map with positive and negative relationships (Carley, 1993). Carley used computer software to determine concepts, and constructed them into maps by hand (Carley, 1993). At this point, Carley suggested that this could be done by both computers or hand, and if done by computers, that the model should be a simple one, and that the more complicated the computer model became the less reliable it would potentially be.

Wise and others had inklings that visualizations of text could have the ability to convey information to the reader much faster than reading the documents themselves (Wise et al., 1995). Evidence to formalize this notion has later been documented using visualizations of topics within a document to increase comprehension (Broek, 1995; Majooni et al., 2015). The importance of this application is most evident when text documents (e.g. entire databases) could not be fully read by a reader. Wise notes that natural language has semantic value that goes beyond frequencies of words within a document and analysis of combinations of words is necessary to determine true topics and themes. Wise used what he calls “Galaxies”, now referred to as clusters, to arrange his topics in two-dimensional space to show how closely connected different concepts are within his analysis of groups of documents, such as an entire weeks’ worth of CNN news stories (Wiese et al., 1995).

An early explanation and practice of Topic Modeling was used by Lafferty & Blei in 2009 to analyze the archive of the journal Science. This massive corpus included all publications in the journal from 1980 – 2002. Using statistical methods of latent Dirichlet allocation (LDA) developed by Blei in 2003, Lafferty & Blei were able to uncover themes and connections between themes using a statistical algorithm to parse through the corpus (Blei, et al., 2003). LDA is a general statistical model which automatically discovers topics within sentences or groups of sentences. Topics are defined as “a distribution over a fixed vocabulary of terms” (Lafferty & Blei, 2009). This work has served as a basis for Topic Modeling and Content Analysis. Some of the most powerful findings of this research include the ability to discover relationships of topics within a document, as well as between documents. This provides the potential for determining how the relationships between topics change over time.

More recent uses of informatics, the science of processing data for storage and retrieval, have been to examine trends and changes elsewhere. Bender et al. used content analysis methods to analyze the structure of breast cancer groups on Facebook to uncover that fundraising was the primary purpose of the majority of these groups (Bender et al., 2011). Content analysis informatics has also been used to classify suicide notes in an attempt to create a deeper understanding of what is going through the minds of troubled individuals. It discovered that situations, relations, emotional states, and cognitive states were leading topics (Pestian et al., 2010). Another notable example examined tweets on e-cigarettes on Twitter to track the increase in dialogue on e-cigarettes over 2012-2014 as well as concluding that the sentiment surrounding twitter dialogue was typically positive (Cole-Lewis et al., 2015). Content analysis and informatics have a wide range of applications, with policy or planning documents being just a fraction of the potential applications.

## **Previous Efforts Using Informatics on State Water Plans**

Earlier efforts to evaluate water resource management in Texas using topic and content analysis as part of an informatics approach, indicated that informatics approaches are useful for determining policy effectiveness as well as that visualizations have potential for increasing the communication of information (Wehner, 2011). Wehner uses a combination of comma separated value spreadsheets and the Sci2 application as a workflow to analyze evolution of the plans over time (Lind, 2011). Sci2 Is a modular toolset specifically designed for studying science. The methods and tools used in this study implemented a different set of workflows and used a combination of open source libraries and tools, such as statistical packages in R and Texttexture with Gephi, as well as commercial content analysis and network modeling software Leximancer (what is R?, 2017; Smith & Humpherys, 2006; Bastian & Heymann, 2009). The tools of R, Gephi, and Leximancer provide more powerful insights, with statistically significant information that goes beyond the power of Excel and has much greater visualization effectiveness than Sci2. This research also begins to explore the regional water plans and how they are connected with the state water plans.

## **CHAPTER IV – METHODOLOGY**

Multiple methods were employed to identify concepts encoded in the corpus of the state and regional plans, starting with straightforward approaches and progressively increasing methodological complexity. Initially, each text was analyzed using frequency counts and advancing through a variety of coding and software workflows that included libraries from the language R, Leximancer Topic Modeling Software, and Texttexture to visualize texts as networks. In order of increasing complexity, after preprocessing, the flow of analysis is as follows:

1. Frequency Analysis – sorting the document(s) individual words by frequency and comparing these across time frames
2. WordCloud Generation – using the frequency charts to create visual qualitative representations of this data.
3. Correlation Charts – using RStudio to identify particular topics from the frequency analysis to analyze what are the most correlated words and understand why they are correlated to one another.
4. Change of Indexed Pervasiveness Over Time – analyze the change of pervasiveness of particular topics between State Water Plans to understand how discussion on certain topics has varied throughout time.
5. Content Network Generation – visualize networks of topics within each plan, understand what differences occur between plans in both terms of space (Regional Plans) and time (State Plans).

### **Preprocessing and Preparing the Dataset for Analysis**

Once documents have been selected, the dataset (corpus) must be prepared for analysis. The preprocessing and data preparation are crucial to completing accurate and

reliable evaluations. Principally, the dataset must be cleaned, tokenized and stemmed. Common practice for text mining in general includes standardizing and normalizing the documents. For example, the algorithms must identify each word so that they can be categorized correctly in later steps. To aid this process the data is ‘tokenized’ and each word or individual data element is separated from others so that it can be easily recognized. Frequently tokenization is completed using clear characters such as “,”, “.”, or “}”. Additionally, texts are scrubbed free of 'stop words': these are common words that don't add value and distinction to the meaning of terms. Stop words include common prepositions, verbs, and pronouns which help bridge ideas within a document together, but don't intrinsically add value to the document. The documents are stemmed, removing these words, removing punctuations, numbers, and also dropping the s from the end of words in plural states. Once preprocessing is completed the first analytical approaches can be applied.

## **Word Counts and Frequency Analyses**

To begin a text analyses of the plans, word frequencies were completed and presented as both histograms and WordClouds (fig 6 and Appendix A). Each document was analyzed to capture the most commonly used words in each plan and the relationships between words, as well as look at how both the vocabulary as well as diversity of words has evolved over the past half century. Initial results were depicted graphically as histograms, which show the distributions of most frequent words, and visualizations of the same counts as WordClouds. WordClouds in each figure are qualitative visual representations of both word frequency, as well as the frequency of each word relative to the size of the entire document. Although they are not meant to



## Topic Modeling in Rstudio

While frequency counts enable a simple straightforward approach, additional insight can be achieved using more advanced techniques such as topic modeling within a corpus. Topic Modeling is a form of natural language processing which is based upon a statistical model used to extract abstract themes and fundamental ideas from the corpus; where a corpus is a collection of texts. For this study, Rstudio and the R programming language was used to implement an initial workflow for topic modeling.

Rstudio, is a graphical interface for using the programming language R. “R is a language and environment for statistical computing and graphics” (What is R?, 2017). This language is a ‘GNU’s Not Unix!’ (GNU) project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues (What is R?, 2017). This language has libraries to perform statistical analyses, while also allowing for packages to be developed by users to further enhance the usability and efficacy of the application. For the bulk of this study the packages *PDFtools* and *tm* were primarily used for analyses. The *PDFtools* package was developed in 2016 and serves as a basis for manipulating pdf documents (Ooms, 2016). The *tm* package developed by Ingo Feinerer in 2015 is used as a medium for the topic modeling method. The full infrastructure of the package was originally published within the Journal of Statistical Software (Feinerer et al. 2008). The *Pdftools* package was used to manipulate the TWDB Water Plans to preprocess the corpus prior to subsequent use with the *tm* package. Once text files were created, white space was stripped, stop words were removed and the documents were ready for analysis.

Using the topic modeling package with the corpus of all the state water plans from 1961 – 2017 allowed various relationships between themes to be revealed. All words, or topics, are correlated with each other to varying degrees, and can be used to create a visual network of connections. Some topics such as drought, were closely correlated with a large number of other topics, such that the correlation limit had to be reset to a high value of 0.98. Other topics such as water were able to be analyzed with a correlation limit of 0.9 to get results with a similar number of connected topics.



Figure 6 indicates the statistical correlations of water with other topics. Correlations between words indicate that those topics frequently appear near each other within the text. Groups of words (or topics) which are all highly correlated with each other indicate a community of topics. This is further explored with the use of Gephi to analyze network composition as laid out in Table 1. The value of these correlation lists given by R is a closer look at the communities surrounding particular topics. This gives a close-up snapshot of topics which are connected to each other in a list view. The networks which are derived from these documents are structured based on the correlations between topics, and provide visual representations of the lists which are generated here.

The charts below show the themes most connected to water with a correlation of 0.9 or higher. “Districts” was the most correlated, with “conservation”, “demands”, and “supply” being some of the more interesting connections that closely followed. It takes a visual approach to make sense of this chart and some knowledge of the TWDB and the State of Texas as some of the subjects which are correlated do not make sense. The first topic “districts”, and fourth topic “conservation” refer to Groundwater Conservation Districts which were discussed earlier. The next two topics, “five” and “percent” do not make much sense. More than likely these two are connected repeatedly through the corpus “five percent...”. Investigation into what topics correlate with “five” and “percent” may have yielded insight. “Demands”, “population”, “producing” [wells], indicate that the TWDB is dedicated to focus water planning on both agriculture as well as municipalities. The equal levels of correlation (0.96) on these two topics reveal that they have carried similar if not equal importance to the TWDB through the past half century of water planning.

\$water									
districts	five	percent	conservation	demands	followed	number	population	producing	
0.98	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.96	
315	increase	inputs	less	quickly	sectors	six	supply	weather	
0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	
180000	additional	consistency	dockum	drinking	dry	groundwater	largest	projected	
0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	
reductions	significantly	well	386	alan	blossom	capitan	early	henry	
0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.93	0.93	

Figure 6: continued next page.

Figure 6. Words Correlated with Water in the Cumulative State Water Plans

Figure 7 below shows the most correlated themes, with a limit of 0.978 correlation. Due to the package having a significant figure limit of 3, this correlation limit defaulted to 0.98. “Desired”, “quick”, and “record” were the topics that best correlated with drought. Correlation of “abnormally”

```
> findAssocs(corpus, "drought", corlimit = 0.978)
$drought
```

desired	quick	record	abnormally	acknowledgments	adams	administratively
0.99	0.99	0.99	0.98	0.98	0.98	0.98
albert	alva	ammerman	andruss	approves	archuleta	ask
0.98	0.98	0.98	0.98	0.98	0.98	0.98
atom	bailiff	balin	ballou	barbara	barho	barrera
0.98	0.98	0.98	0.98	0.98	0.98	0.98
bartos	baucum	baumgardner	beard	bearden	beef	bennett
0.98	0.98	0.98	0.98	0.98	0.98	0.98

Figure 7. Words Correlated with Drought in the Cumulative State Water Plans

Another informative example is the topic of pollution. “Acceptability”, “governments”, “recycling”, and “spills” were among the most connected topics with a correlation limit of 0.9 or higher.

```
> findAssocs(corpus, "pollution", corlimit = 0.9)
$pollution
```

amended	governments	acceptability	presidents	proofing	floodplain	initiate	capability	recycling	spills
0.95	0.93	0.92	0.92	0.92	0.91	0.91	0.90	0.90	0.90

Figure 8. Words Correlated with Pollution in the Cumulative State Water Plans

## Indexed Progression of Common Topics

Another technique used to analyze the corpus is to determine how the prevalence of certain topics change over time. This idea has been used to visualize how language regarding nuclear proliferation changes within corpus’ over the course of the Cold War

period (Jacob et al., 2015). It also shows up in analyses on how language and sentiment regarding Muslims in social media discourse within Swedish online forums has changed over time (Törnberg &Törnberg, 2016). Taking topics, and comparing their prevalence within a corpus over time is yet another powerful and novel method of explaining and seeing how trends are unfolding.

Using the 1961 – 2017 state water plans as the corpus a handful of emerging topics (plan, strategies, needs, environment, drought, future, reservoir) were compare across all 10 plans. The frequency of each word was divided by the total number of words within its respective document to create an index of pervasiveness. The pervasiveness index was then graphed to show how it evolves through the past 56 years. of water planning.

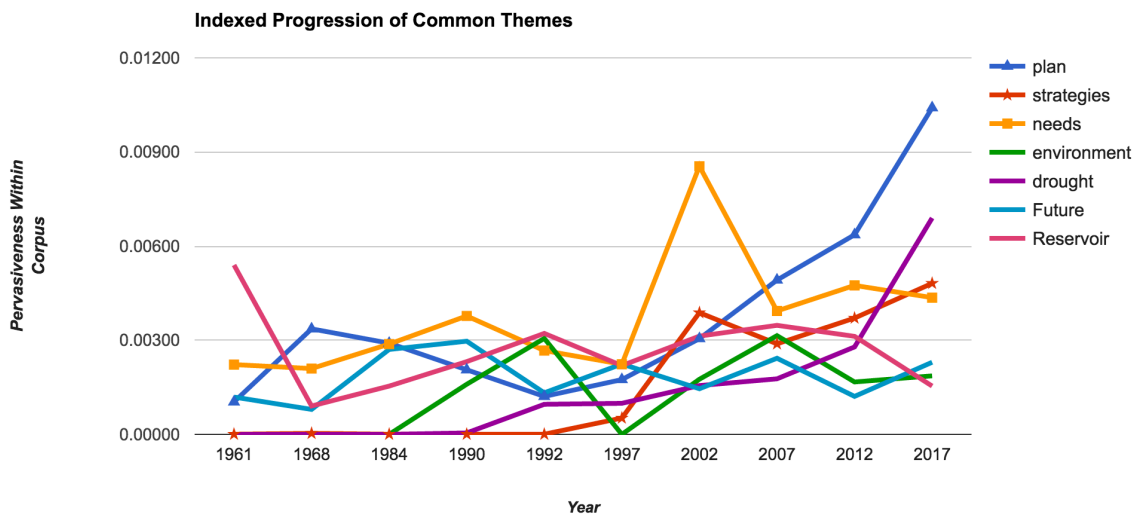


Figure 9. Change in Pervasiveness of Common Themes Across all State Water Plans

The ‘plan’ topic’s pervasiveness held stable until the 1990’s when it began trending upwards. “Environment” and “strategies” were both nearly nonexistent in the

earlier plans, and have gained traction in the recent few decades. “Drought” has also seen a relative uptick in recent decades.

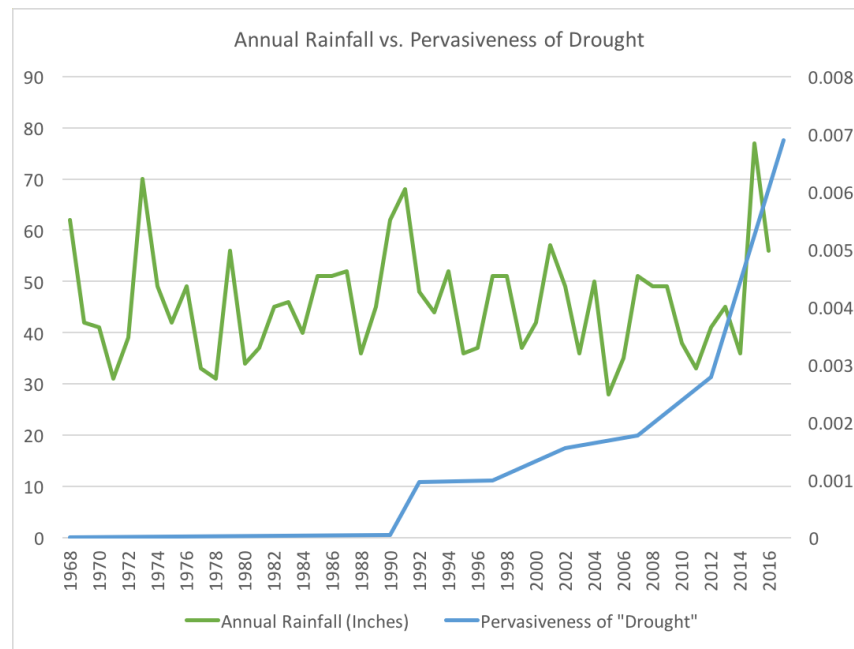


Figure 10. Annual Rainfall vs. Pervasiveness of Drought in the State Water Plans (rainfall data from Lloyd, 2017)

The increased discussion on “drought” most likely has to do with the occurrence of droughts in Texas. 2011 was a year of low precipitation in the state, and this coincides with the rise of discussion on drought which increases most in the 2012 and 2017 State Water Plans. While this makes sense for the recent drought, it doesn’t align with the previous droughts of Texas. Texas is known for having highly variable rainfall with periods of drought, followed by periods of flood. To try and visualize this, Figure 10 displays a line graph which overlays both annual rainfall with the pervasiveness of “drought” in the State Water Plans. While the pervasiveness only increases in recent time, precipitation data shows many dry and wet times, with the 2011 rainfall record not being significantly worse than previous droughts through history. This suggests that the

pervasiveness of “drought” has no correlation with rainfall. The data for the state rainfall records was taken from Texas A&M Agrilife records (Fig. 11), (Loyd, 2017).

Examining other datasets, it turns out the pattern of “drought” more closely follows the population growth of Texas (Fig. 10). This correlation does not begin to occur until the state population reaches a particular point, about 17 million. Once this population threshold was reached, the impact of drought on populations has become significant enough to warrant increasing discussion on “drought” in the State Water Plan. This insight suggests that the discussion of drought is primarily based on municipal water supplies. Texas cities typically rely on reservoirs to serve the population. Reservoirs are fed mostly by surface water, and are strongly impacted by droughts. It is expected that “drought” will continue to rise as an important topic in the State Water Plans as long as the state population continues rising. The data for the state population was taken from the U.S. Census (Fig. 10), (U.S. Census, 2016).

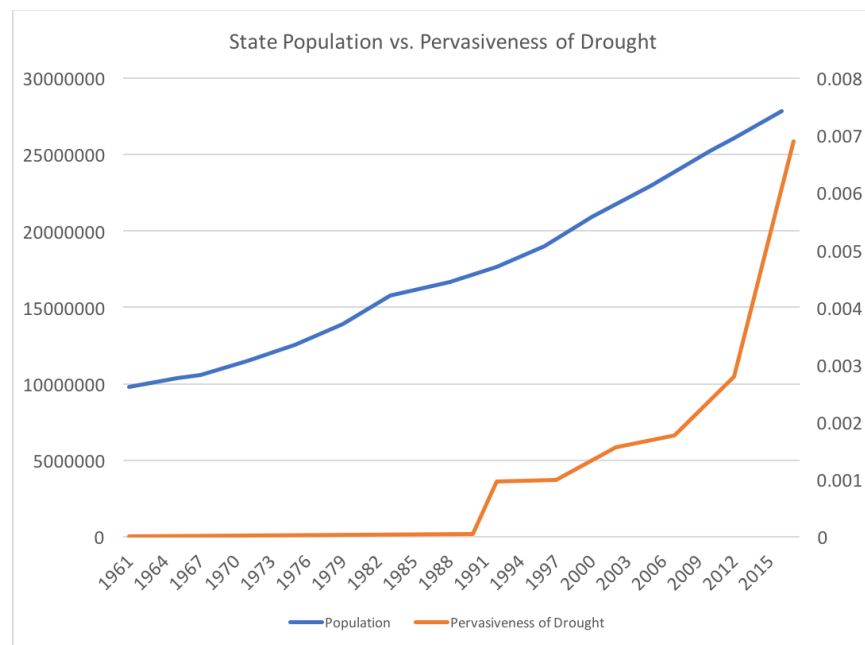


Figure 11. State Population vs. Drought in the State Water Plans (state population data from 2016 U.S. Census)

### **Commercial Concept Mapping Tool: Leximancer**

Another method used to analyze the corpus for this study was the software Leximancer. Leximancer is based upon semantic and relational methods (Smith & Humpherys, 2006). Using statistical techniques of latent semantic analysis (LSA), Leximancer works to group topics together into themes and create concept maps. The software provides a network visualization of the corpus. The advantage of this is to create a statistically significant visual representation of themes. The visualization is similar to WordClouds because users are able to quickly look at a concept map and retrieve the main concepts of a corpus. The techniques to develop each concept map are more mature and more likely to identify unexpected insights. Concatenated with the evidence that visualizations increase comprehension documented by (Majooni et al. in 2015), it is expected that these visualizations increase comprehension of the documents to readers.

Using the most recent 2016 regional water plans from the TWDB in Leximancer, concept maps were developed. The settings used in the Leximancer software package to produce figures 12 and 13 are located in Appendix F. The Brazos Region G was not able to be analyzed by the Leximancer due to the fact that the PDF copy that is available is a photocopy of the plan, and not a textual copy like the rest of the plans. Because of this unfortunate incongruence, this region was not able to be compared with the rest of the Regional Water Plans of 2016.

Below is an example of the concepts within the regional water plans, comparing a region in West Texas with one in East Texas. Since Texas has a hydrological gradient of arid conditions in the west and humid wet conditions in the east, these two regions are expected to have contrasting needs, thus were chosen for further discussion in this document. A full set of graphics for state water plans over time and regional water plans are included in Appendix B.

Leximancer denotes more important themes with warm colors (red and orange) with less central themes as cooler colors (blue and green).

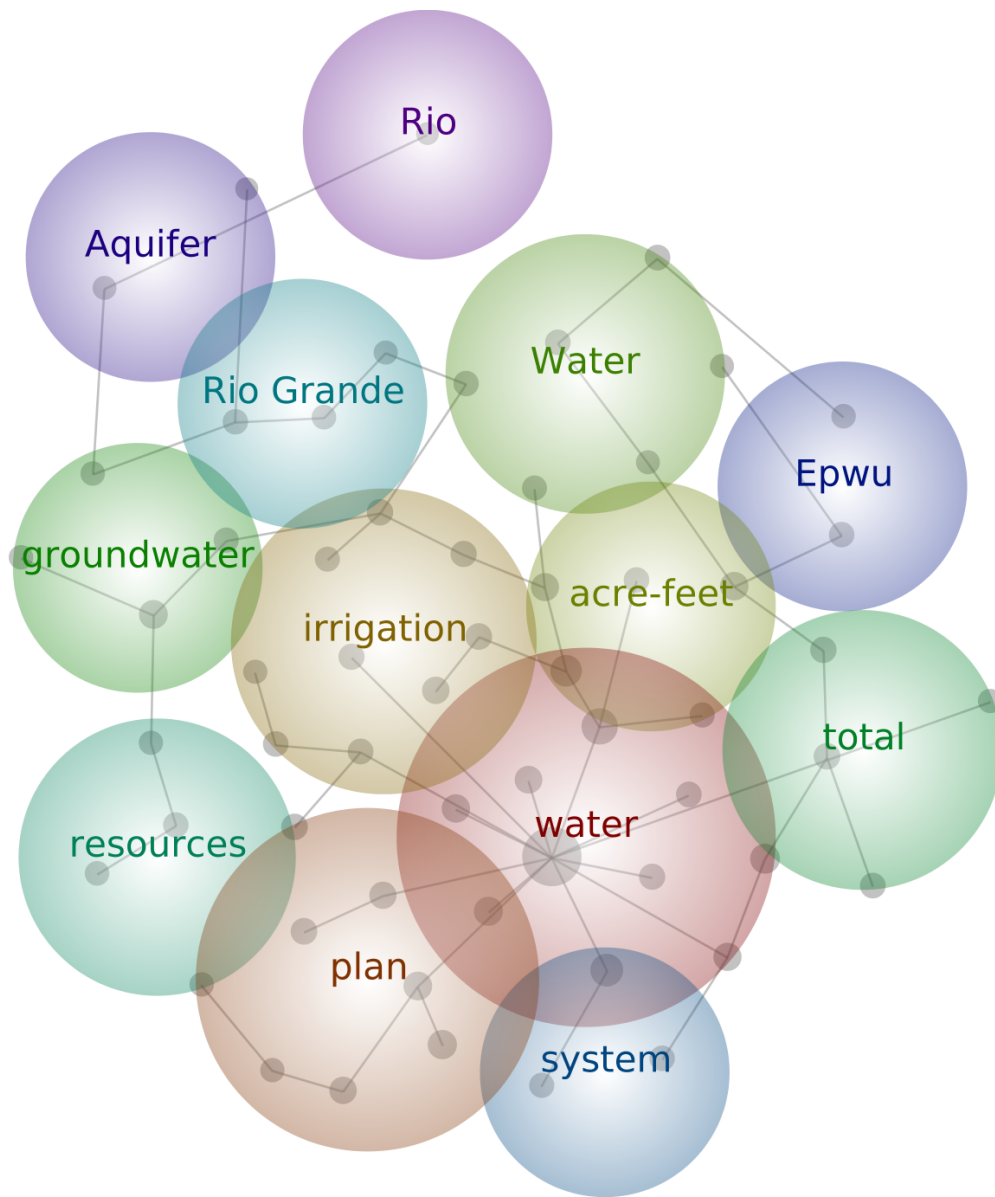


Figure 12. Leximancer Network Visualization of Region E (West Texas) Regional Water Plan

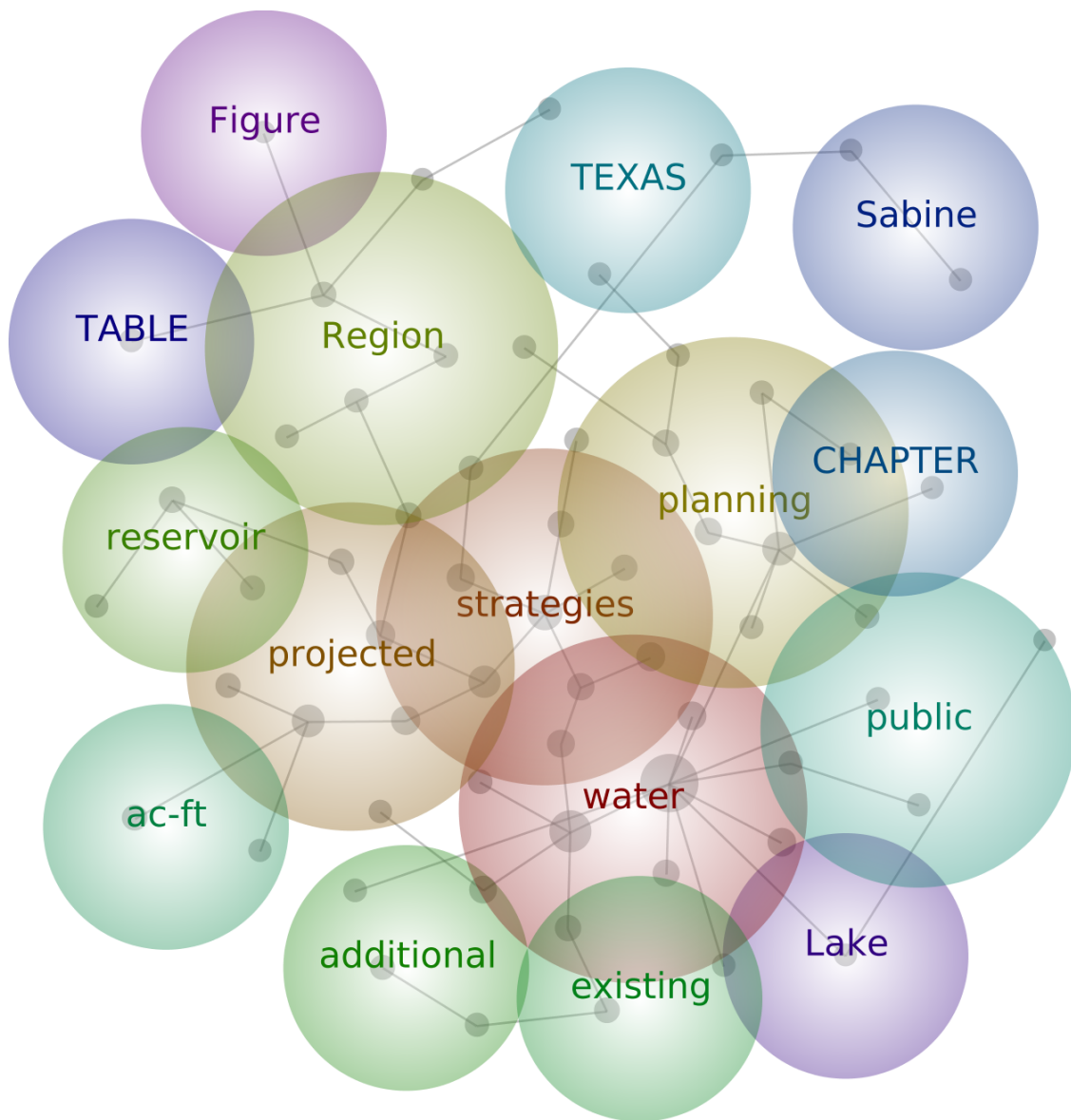


Figure 13. Leximancer Network Visualization of Region D (East Texas) Regional Water Plan

Water is of course the central theme of both regions. The concepts of both regions have more commonalities than differences. Plans and groundwater are some of the other commonalities between the regions. The major difference is “irrigation” appearing in the West Texas concept map, compared with “reservoir” and “public” appearing in the East



Texas concept map. Region E in West Texas has a more sparse population than Region D in East Texas. The arid nature of West Texas correlates with the need for irrigation plans, which is absent from the concept map in the wet geographical location of Region D in East Texas. East Texas' population is more dense and growing faster than the population of West Texas which is more consistent with the emergence of concepts such as "reservoir" and "lake". The focus of East Texas water planners is based upon the need for surface water to be used for municipalities, and human consumption. This gives insight into the primary goals for both regions, which correlate with what is already known about the regions' geographical and sociometric relationships.

Region E within the Far West Texas Region has an emergent topic of "Rio Grande" which is exclusive to this regions Content Analysis (Appendix A). Although the Plateau Region J, and the Rio Grande Region M both also lie along the Rio Grande River, this theme does not readily emerge from the Content Analysis. Far West Texas is sparsely populated apart from El Paso. This region has historically been dominated by agriculture and has not had major population growth relative to other parts of the state. The sources of water for this agriculture comes from both groundwater as well as canals from the Rio Grande River, most prominently the 31 mile Franklin Canal which is located near El Paso and developed in 1889 (Powell, 1917). Revealed by a Community Environmental Scan by the McAllen Chamber of Commerce, the predominant economic structure of McAllen, and thus the Rio Grande Valley is based upon international trade, retail and tourism (McAllen, 2013). While McAllen and the Rio Grande valley has been historically an agricultural based community, it has transitioned away from this as the population has grown substantially. This is well aligned with the planning for that region. The 2016 Regional Plan for the Rio Grande Region M indicates one particular theme of "treatment", referring to water treatment for tap water. Although multiple regions lie along the Rio Grande River, the initially surprising find that the 'Rio Grande' Region does not have that endemic term emerge, is explained by the different economic circumstances of each region.

Of the major rivers of Texas, the Sabine, Rio Grande, Colorado, Brazos, and Guadalupe are emergent themes from the Regional Water Plans. The Red, Trinity, and Nueces rivers are ones which are surprisingly left out of these concept maps. The Red Rivers' absence can be explained due to the fact that it is bordering the northern boundary of Texas and Oklahoma, coupled with the large population growth of those regions (B, C, and D) where the regions' water needs are more focused on providing reservoir storage for the municipalities. Region H, has the emergence of "Brazos", and absence of "Trinity" in its concept map. The Brazos river is much larger, explaining its appearance within the concept maps. Similarly, the larger Guadalupe River emerges from the concept map of the South Central Texas L region, and the smaller Nueces River is not as high priority.

An unexpected finding from the Regional Water Planning concept maps was the "reservoir" theme in the Llano Estacado Region O. The reservoir theme is common in North and East Texas, where populations are growing at faster rates than West Texas. The passing of the Texas Senate Bill 675 in 2007 designated a site in the region to be an additional lake to the Canyon Lake System to supply more water for municipal use. In depth look into the Region O plan reveals the discussion of this potential future reservoir site. As of 2015, 3 of the 4 reservoirs which supply water to Region O were less than 10% full (LERWPG, 2015). Although the region is not known for its large population, the imminent danger of these unusually low lake levels and subsequently extensive discussion on the regions' plan explains why this theme emerges within this region, and is absent from the top themes of other regions in West Texas.

## **Topic modeling in networks with Texttexture and Gephi**

Texttexture is another online software tool used to analyze topics and visualize text. Self-dubbed “The Non-Linear Reading Machine” Texttexture’s provides an easy interface for analyzing text documents and implements recognized topic analysis methods (Nodus Labs, 2012). Just as in the previous methods for visualizing the corpus, Texttexture removes stopwords and co-occurrences of words such as ‘took’ and ‘take’ are converted to ‘take’. Paragraph and sentence structure are taken into account as well. The resulting correlations between words, are used to create nodes with linkages between the topics. The results are encoded in a network format with edge and node information that can be visualized using Gephi Java toolkit, an open source graph visualization program (Bastian & Heymann, 2009). Size and distance of the nodes is a representation of “their betweenness centrality; that is a measure of how often a node appears on the shortest path between any two randomly chosen nodes in the network” (Nodus Labs, 2012). Finally, the nodes are colored based upon which community they belong to (Nodus Labs, 2012). While the Leximancer concept maps are color coded based upon the relative importance of topics, the Gephi visualizations do not follow this pattern. The communities are arranged by color but warmer colors do not represent more important communities.

Figures 11 and 12 provide a few examples of the networks created from the state water plans of 1961 and 2002. Additional graphics are included in Appendix C. Texttexture creates an interactive graphical result, which is best viewed within a web browser.

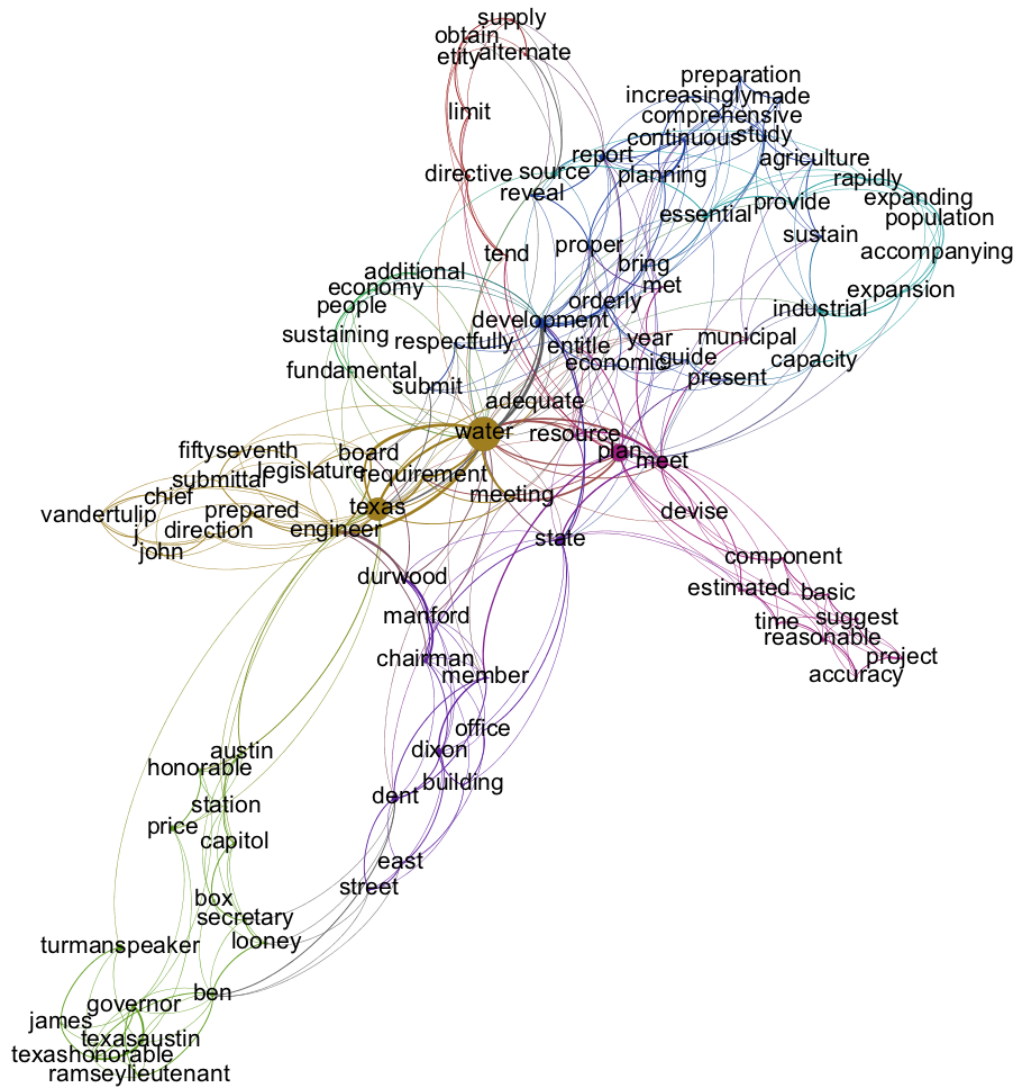


Figure 14. 1961 State Water Plan Network Visualization - Visualized with Gephi

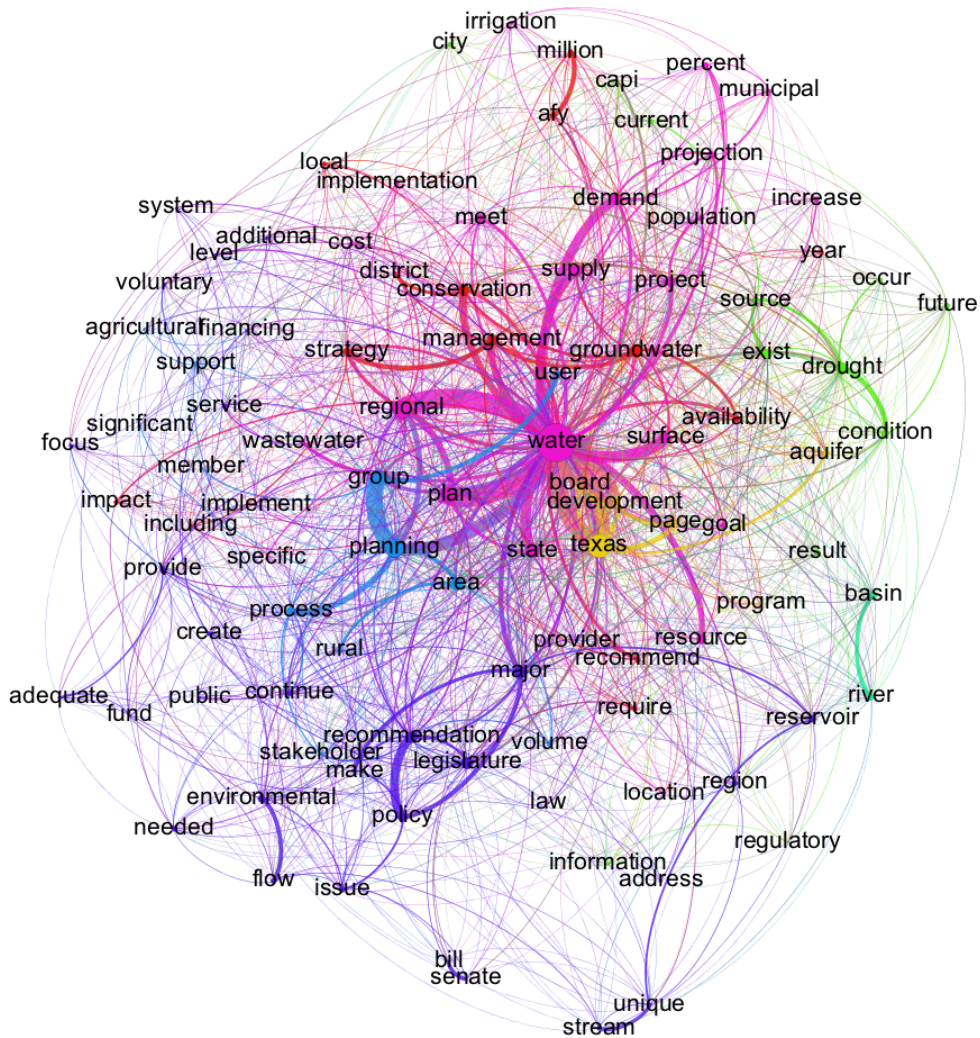


Figure 15. 2002 State Water Plan Visualization – Visualized with Gephi

In both cases, all roads lead to water. In fact, all edges lead to water; all State Water Plans have the central node of water (Appendix C). Gephi organizes communities of topics by color (Bastian, 2009). Although entangled in figure 12, the 1961 network in figure 14 better organizes into communities, which is a reflection of the wide diameter of the network. A wide diameter of the topical network is indicative that the themes discussed within the plan are less related, and less connected to one another. These two

networks highlight large structural differences between the two documents while still having many commonalities, primarily the desire to meet the needs of Texans for tomorrow.

Beginning with the 1961 State Water Plan in Figure 14, the green portion of the network is grouped together in a community of state politicians. “Governor”, “Austin”, “Honorable”, and names such as “Looney”, “Turman”, and “Price” refer to the State Legislation. Slightly above and to the left, the brown community in this network yields topics of “Prepared” and names “Vandertulip” and “John” which are authors of this State Water Plan. The Content Analysis algorithm groups these communities independent of each other, but next to each other relative to more distant communities. They both connect to the intermediate node “Texas” before connecting back to the source “water”. On the other side of the network there is another secondary node “development” which unites three more distinct communities; the red, blue, and turquoise communities. The red community is a group of topics based upon “supply” and “obtain[ing]”. The blue community next to this is a collection of “preparation”, “increasingly”, “continuous” topics which is a cluster of words used to describe other topics, and are not actually beneficial for analysis. The aqua community, “expansion”, “population”, “rapidly”, and “industrial” points to the essence of why the TWDB was developed and why the State Water Plans were necessary. Post World War II growth across the country, increases in industry, and overall expansion accumulated to the point that the TWDB was created, and is illustrated here in this aqua community. A final secondary / intermediate node connecting the pink community to “water” is “plan”. The topics in the community are comprised of “project[ed]”, “estimated”, “component”, and “accuracy”. This community is likely formed by the timeline and future projection of the goals within this initial water plan. In summary, the 1961 State Water Plans major communities are important figures and names relevant to the TWDB and state government on one side, with projections, plans, and subjects (economy & citizens) comprising the other half.

The 2002 State Water Plan was chosen for a deep in-depth discussion due to its particularly contrasting behavior with the original state plan (Figure 15). This plan has

the smallest diameter of all of the State Water Plan networks, while the 1961 State Water plan has the largest diameter (Table 1). The communities are more intertwined and overlap significantly more than in the 1961 plan. This 2002 plan does not have discernible intermediate nodes which connect communities to the central node “water” as was described in the discussion on the 1961 plan. There are three communities that stand out within the 2002 State Water Plan; green, pink, and purple. The green community located on the far right side of the network is filled with hydrologic terms “drought”, “aquifer”, “basin”, “river”. The pink community refers to “population”, “municipal”, “supply”, and “demand”. The largest community which seems to encompass at least half of the entire network is represented by “environmental”, “impact”, and “agriculture”, as well as “financing”, “fund”, “public”. While the green and pink communities first described seem to have major commonalities within the communities, hydrologic terms and people respectively, this final large community has many topics which don’t seem to work well together. An explanation of this is that the 2002 State Water Plan has more themes that overlap into multiple communities, and make the topics appear to be more integrated. An alternative take is that the 2002 plan is more recursive, meaning that the plan refers to itself multiple times. Evidence to support this is in table 1 in the edges column. The number of edges in the 2002 plan is four times greater than the 1961 plan. The topics/nodes within this plan have greater connectivity to the other nodes.

Another approach to analyzing these networks uses the measure of centrality. A topic or node that is more central to the network indicates a higher level of importance in the plan. For instance, in the 2002 State Water Plan network the topics “surface” [water] and “groundwater” are near each other, and relatively equidistant from the center node of “water”. This indicates the equal valuation of both surface water and groundwater by the TWDB and the State Water Plan. Comparing two more nodes “stream” and “river” it is apparent that “river” is closer to the central node as would be expected the TWDB puts more emphasis on rivers than streams.

## **Statistical Analysis of State Water Plans Topic Networks**

The networks for the state water plans were initially created and visualized in Texttexture. The gexf source code for these networks was then downloaded and used in Gephi Software. Not only is Gephi an excellent tool for manipulating visualized networks, but it also has statistical analysis tools that are useful for uncovering network structure, as well as comparing networks against each other. Using the modularity algorithm, Gephi is able to calculate the number of communities within a network, as well as the modularity measure (Blondel, et al., 2008). The modularity of the network is a measure of how close or distant these communities are to each other (Blondel, et al., 2008). A higher value of modularity represents more distant community clusters within the network. A community within a network such as the state water plans represents a group of topics which are closely linked. Diameter is defined as the shortest distance between the two most distant nodes in the network. Average degree is a measure of the average amount of edges coming out of all nodes. Average Path Length is a measure of networks that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes; that is a smaller measure of Average Path Length represents a smaller ‘distance’ between all concepts within the network. For simplicity, all networks derived in Texttexture used a standard of 100 nodes. The number of edges connecting these nodes vary, as well as the diameter of the network, the average path length, average degree, modularity, and number of communities. The table below illustrates that the values for all of these different metrics vary from plan to plan (Table 1).

The state water plans have five to eight communities defined by the Gephi algorithms. The number of communities determined by Gephi contrasts with the number of communities that can be distinguished by qualitative visual assessments of the networks described in the previous section. Additionally, the modularity values range from 0.17 to 0.54 with the largest range for this measure happening between the first two documents. Average degrees range from 8.1 to 31.7. Average Path length has a range



between 1.9 and 3.9. Diameter of the networks had a large range of 3 to 10 nodes. The edges of the networks varied between 405 to 1701.

	<b>Nodes</b>	<b>Edges</b>	<b>Diameter</b>	<b>Average Path Length</b>	<b>Average Degree</b>	<b>Modularity</b>	<b>Number of Communities</b>
<b>1961</b>	100	405	10	3.919	8.1	0.54	8
<b>1968</b>	100	1105	4	2.02	22.1	0.17	6
<b>1984</b>	100	1584	3	1.89	31.7	0.2	6
<b>1990</b>	100	1339	4	1.99	26.8	0.255	5
<b>1992</b>	100	528	6	2.84	10.5	0.4	6
<b>1997</b>	100	1181	5	2.04	23.6	0.22	5
<b>2002</b>	100	1701	3	1.88	34	0.22	6
<b>2007</b>	100	1546	3	1.91	30.9	0.22	8
<b>2012</b>	100	1655	4	1.9	33	0.225	8
<b>2017</b>	100	1414	6	2.3	28.3	0.28	6

Table 1. TWDB State Water Plans Statistical Analysis of Generated Networks

## **Regional Water Plans Represented by the State Plan**

Since the State Water Plans' goal is to represent the 16 state regions and be a sponsor of their individual plans, a final analysis is to compare the networks of the 2016 Regional Water Plans to the 2017 State Water Plan. One way to do this would be to collate all 16 regional water plans into one text document, and run it through Leximancer or Texttexture to compare the network with the State Water Plan network. The drawbacks of this would be that not all regional plans are of equal length, not all regions are of equal size, population, or economic output. This means the resulting collated regional plans' network wouldn't accurately represent the state. Another way to compare is to visually compare the 16 regional plans with the single state plan to look for commonalities.

Using this approach, a first observation is that “water” is the central node for all regional plans as well as the state plan. Comparing the community topics of each region with the State Water Plan, there are topics which appear in both as well as topics occurring in regional plans which are missing from the state plan. “Strategy”, “cost”, “plan”, “supply”, “irrigation”, “aquifer”, “groundwater”, and “plan” are commonalities between both sets. Topics which are missing from the State Water Plan but occur in the Regional Water Plans typically include regional geographical figures such as “Rio Grande”, “Brazos”, “Plateau”, and “Colorado”. Non-geographical figures which are absent from the state plan include “wastewater”, “reservoir”, and “projected population”. Although these are missing, other similar topics do appear, for instance “municipal” occurs and is correlated with “reservoirs” as well as “projected populations”. The regional plans content networks which were used for this analysis are located in Appendix B, and the 2017 State Water Plan network is located in appendix A.

## **CHAPTER V – DISCUSSION**

### **State Water Plans**

Multi-method analyses render numerous results for consideration in this study. The frequency counts presented in histogram and WordCloud graphics (see fig. 5 and Appendix A) provide a preliminary view of natural segmentation in the water resource terminology. In the case of the first state water plan (1961), the terms naturally fit into 7 buckets or bins in the histogram and different sized and colored groups of words in the WordCloud. The diversity of terms in each successive plan decreased from 5 groupings in 1968, followed by 4 groupings in 1984, 3 groupings in 1990, 4 from 1992 through 2007, and then dropping to 3 in the most recent water plans in 2012 and 2017. Figures documenting each plan can be viewed in Appendix A.

Correlations results generated using the R programming libraries highlighted interesting insights about the relationships among key terms. Of particular interest were correlations between “drought” with “desired”, “quick”, and “record”. It’s possible that desired is from mentions of the state mandated DFCs for every region and “record” is most likely related to the mandated use of the “drought of record” to assess performance and planning measures. Interestingly, ‘quick’ in relation to “drought” may confirm an intuitive interpretation that responses in the event of drought should be rapid. Further examination of the correlation values in conjunction with inspection of the text sections that contain these terms could provide greater understanding. Additional terms, such as pollution, and their correlated community of words merit similar evaluation and further study.

Compared with other uses of Topic Modeling, the volume of corpus that were used in this research is relatively small. This scope of work is known as a micro scientometric study (Lind, 2011). Törnberg & Törnberg (2016) used entire volumes of internet forums to analyze how sentiment towards Muslims in Sweden changed over

time. Other uses of these approaches involve analyzing daily newspapers over the course of decades, etc.

While the TWDB state water plans encompass 603,438 words over the entire corpus, the content remains orders of magnitude lower than macro-scale Topic Modeling experiments. That being said, this research qualifies as a micro-scale analysis and the impact of visualization of over half a million words tells the story of water management evolution at a glance.

Looking at the networks created by Texttexture and Gephi, as well as the community segregation that Gephi suggests, there are common communities which are evident and can be extracted by using these results. The algorithms used in Gephi suggest between 5 and 8 communities identified. Below are 6 common communities of topics which can be extracted by looking at how the 10 networks organize themselves.

1. Government – state officials and dialogue about the capitol.
2. TWDB – members, writers of the state water plans
3. Texas – people and economy
4. Visions – projections, plans, temporal relevance
5. Agriculture – groundwater, irrigation
6. Municipalities – reservoirs, surface water

The various statistical metrics used to compare changes over the state water plans over time indicate that the largest changes in structure occurred between the first and second plan. Modularity, diameter, average path length, and average degree have much tighter ranges within the successive plans. This could indicate that the initial water plan covered a much wider range of topics and ideas, which is illustrated by the large diameter of the topical network of the 1961 plan.

Another aspect to consider when using a limited number of documents within a corpus is the variable range of words within each document. As shown in Figure 14, the number of terms used in each of the state water plans vary from plan to plan, which may

make inter-comparison between years, such as 1997 and 2007, difficult. A smaller corpus size reduces the resolution of analyses and inferences from the results.

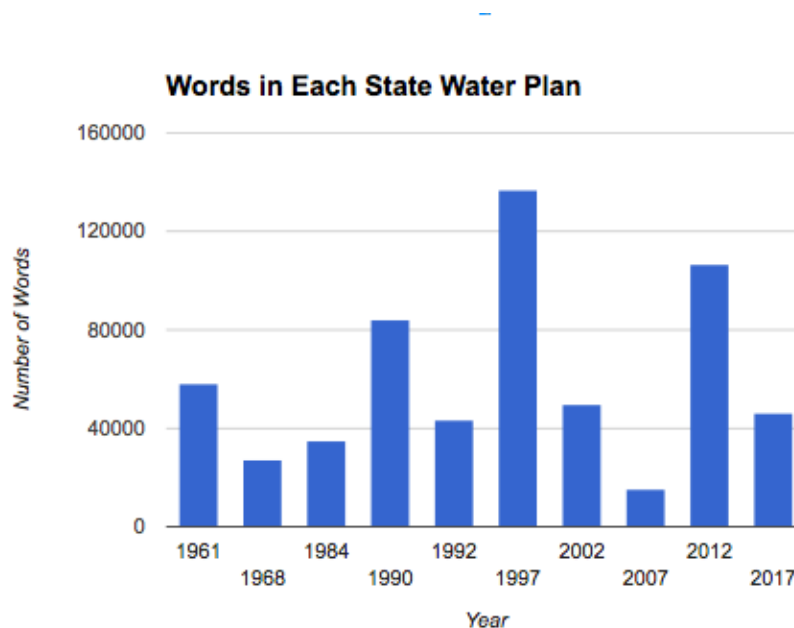


Figure 16. Total Words in Each State Water Plan

Some noticeable outliers on this chart include the two shortest State Water Plans 1968 and 2007. According to the Texas Water Development Boards' site, the 1968 State Water Plans is only available as a summary of the initial plan. There are no remaining digital copies of the full plan. Due to the large nature of the summary version, it is expected to produce similar results through Content Analysis as the original plan would. The 2007 State Water Plan is split into two volumes, with the first volume being the only part of the document used in this study. The second volume is comprised of over 30 sub sections which are each individual documents, and include summaries of the regional plans as well as addenda. This has the potential to hinder the Content Analysis comparisons between the 2007 plan with others. A collation of the individual documents

and reassessment of the Content Analysis for this year is a possibility for further research on these State Water Plans. This discrepancy between the 2007 State Plan document and the others can be used to explain the relative spike in the “needs” group in Figure 15.

## **Regional Water Plans**

Leximancer’s concept maps are a powerful way to increase the understanding of the thousands of pages of documents that have been published by the members of the TWDB, and the regional water planning groups throughout the state. Unfortunately, the commercial software does not openly share the stepwise analyses of the algorithms which make the product useful. Leximancer is therefore limited in its application for further research. Licensing costs also limit the potential use of the product for research.

What can be extracted from the visualizations that Leximancer provides are quick overviews to what the regional water planning groups are interested in for their hydrological futures. It comes as no surprise that water is the central theme for nearly all of the Leximancer topic networks that were generated. These visualizations also give insight into the intrastate variations of needs, west Texas focused on agriculture and groundwater, while municipalities and reservoirs dominate the interests of east Texas. With the comparison of regional networks to the most recent state network, it is apparent that the State Water Plan is an accurate representation of the Regional Water Plans themselves.

## **Reproducible Framework**

One goal of this research was to create a framework for others to use for water policy analysis. Figure 17 is the visualization of this Policy Analysis Framework, which was employed to create the various outputs reported in this study. The R code which was developed for this work is also included in Appendix E.

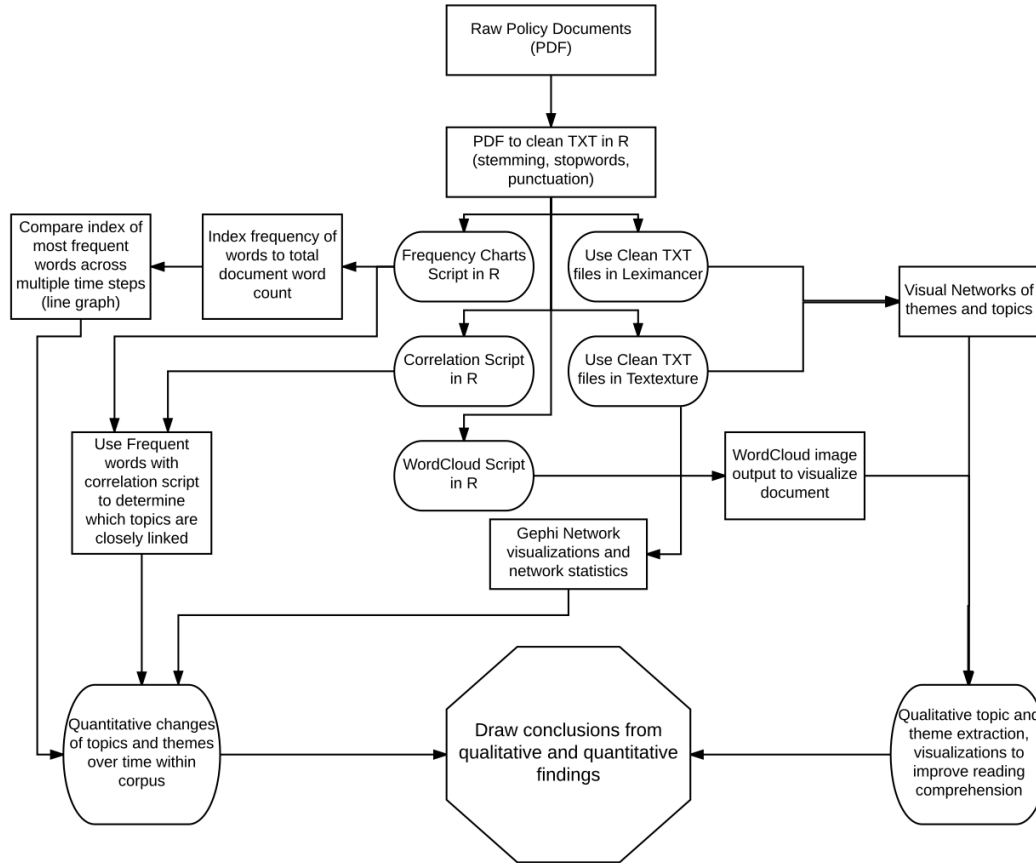


Figure 17. The Policy Analysis Framework for Combining Qualitative and Quantitative Approaches to Analyze Water Policy Documents

The framework produced uses a combination of qualitative and quantitative methods to extract themes and topics, create visualizations of WordClouds and topical networks. Common statistical algorithms are employed in Gephi to compare internetwork structure. Indexed frequent themes are compared across plans using simple algebra and line charts which can be developed in excel. The value of combining both quantitative and qualitative methods has been elaborated on in this discussion. For instance, Gephi is able to statistically determine the number of communities within a document, but that doesn't always match with what is seen on the networks. On top of this, just knowing the number of communities which Gephi tells isn't enough to make policy decisions, a

personal examination of the networks to determine what the communities are composed of and mean are arguably more valuable than the quantitative metrics themselves. Having a historical knowledge of the State, and Regions is also valuable to be able to analyze the content. Appendix E holds the resulting R scripts which were written.



## CHAPTER VI – CONCLUSIONS

Content Analysis is a powerful and proven statistical tool for extracting themes from large volumes of text. This research uses Content Analysis among other methods to analyze Texas state water plans and regional water plans to evaluate patterns in the documents and create visual representations. Results serve a two-fold purpose: to extract hidden topics within the semantics and create visual representations of the topics, which serve as an aid to viewing the content and information in massive documents for both the citizens of Texas and state legislatures. In particular, the research results have potential for aiding policy makers who are responsible for considering recommendations from the TWDB to set budgets, create policy aimed at satisfying the stakeholders, and meeting water resource needs across the state.

The Topic Models and Concept maps produced for this study shed light on how the water resource concerns and issues in Texas change temporally and geographically. East Texas has a higher focus on creating surface water reserves to serve growing urban populations, while West Texas continues to rely on groundwater for irrigating agriculture. Strategies, plans, and environment are topics that have become more prevalent in recent years across the state plans. This ties back in with the desired future conditions table of Groundwater Management Areas in Appendix D which highlights the similarities between the regional planning groups and the GMA's which have jurisdictions that cross each other (fig. 4). The differences between interregional economies is illustrated by the concept maps of the Regional Water Plans as well. The Rio Grande Region concept map is absent of mentioning the Rio Grande River it is named after, primarily due to a shift in its economy away from agriculture and into tourism, international trade, and retail. More surprising emergent topics such as reservoirs in West Texas is explained due to temporally acute lake levels, which raise concern in this plan.

Statistical comparisons between state water plan networks reveals the greatest difference between the first water plan in comparison with subsequent plans. The

network dimensions indicate that the original 1961 plan included a wider range of discussion with fewer overlapping concepts compared with more recent plans. The values for diameter, average degree, modularity, average path length, and number of communities range between any two successive plans is largest between the 1961 and 1968 state water plans. That is, every single statistical metric that was measure within Gephi to compare the networks to one another had the largest range between these two plans. The number of communities between networks varied, but remained between five and eight distinct communities of topics throughout all state plans.

Further work for this research could include a more in depth analysis of the regional water plans and how they compare with the 2017 State Water Plan. Comparing these documents with quantitative data would give information on how accurately the 2017 State Water Plan reflects the interests of the Regional Water Planning Groups. Diving deeper into the most frequent themes of the state water plans, to compare how the pervasiveness of less common topics change over time could also yield valuable information on how the interests of the state and the TWDB change over time. Reaching out to Region G to obtain a textual copy of the plan, or employing text recognition software for this Regional Plan would be useful to bridge the gap of information in this research. This research examines how Regional Plans themes change graphically in 2016, and how the State Water Plans change over time from 1961 to 2017. One way to increase the breadth of analyses would be to analyze individual Regional Plans over time as well. Emergent, temporally acute topics within these Regional Plans will certainly contrast with one another, illustrating the evolution of the needs exclusive to each region. The value in these methods goes beyond historical plans and can be used to understand stakeholder input during the creation of subsequent plans during the period of input and plan design.

State legislators can use the information from this research to quickly grasp the concepts laid out in the State Water Plan by examining the most frequent words and the WordClouds to begin with (fig. 5; Appendix A). The next level of insight comes from examining the network visualizations and how the communities within them reflect the

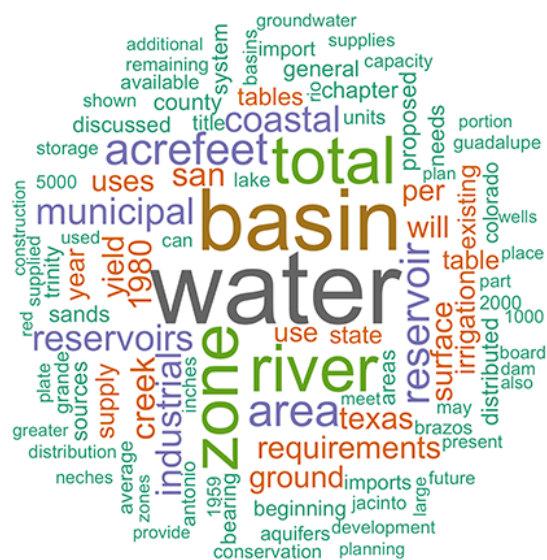
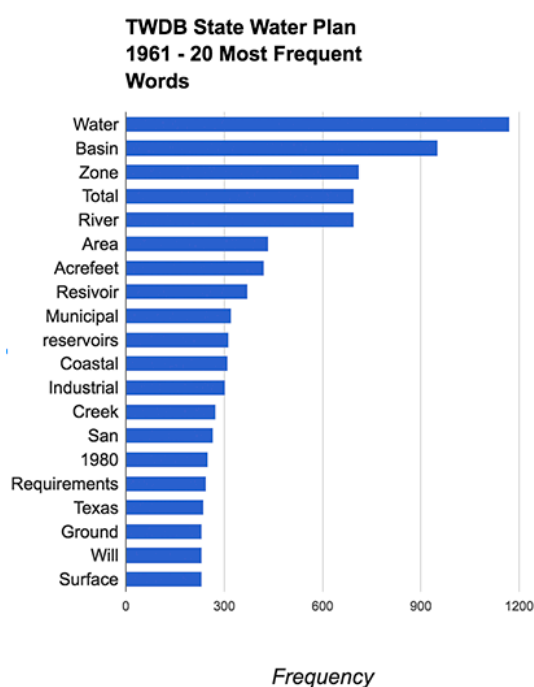
topics in the State Water Plan. These visualizations act as supplementary materials to the State Water Plan itself with the goal of increasing the comprehension of the large documents, to make policy decisions which more accurately reflect the desires of different regions and management areas within the state. While it is unreasonable to expect ordinary citizens to read either Regional, or State Water Plans, the qualitative assessments of the WordClouds, and Concept Maps, and Topic Networks are easily absorbed by the lay person. This creates a way for the authors of the Regional and State Water Plans to open channels of communication with stakeholders within the communities of Texas to determine whether or not the citizens' needs are being discussed and allows for more accurate plans in the future to be developed. One example of takeaways for the State Legislation is that the impact of drought on Texas is primarily dependent on the population.

Visualizing networks of topics is a cross disciplinary, reproducible method for identifying patterns and changes between bodies of texts. This has many useful benefits to scientists as well as policy makers. Some people say that a picture is worth a thousand words, but in the case of Content Analysis Networks, the picture is worth tens of thousands of words and increases the ability to understand patterns and trends that are not readily observed from other approaches.

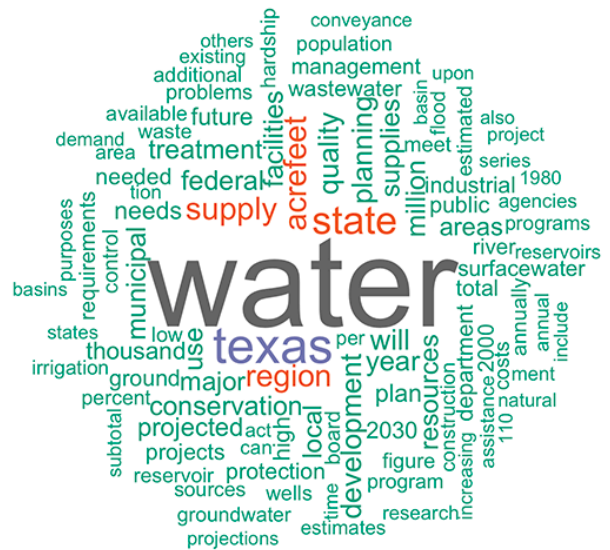
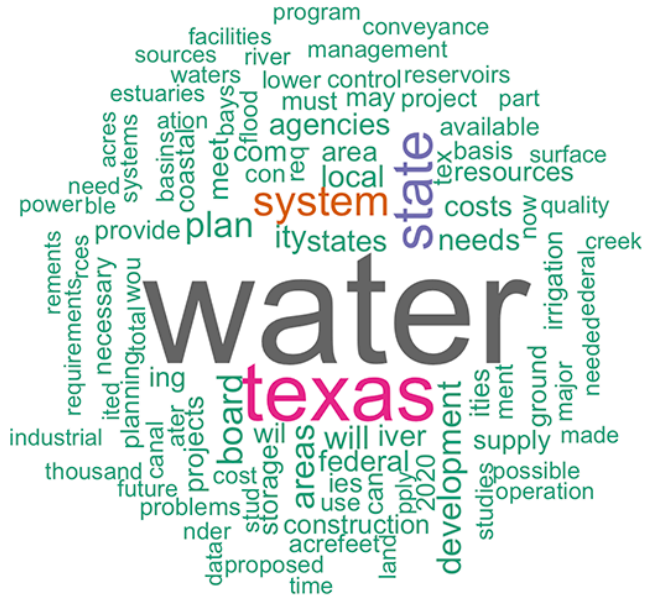
## APPENDICES

## APPENDIX A - FREQUENCY ANALYSIS: HISTOGRAMS AND WORDCLOUDS OF STATE WATER PLANS 1961-2017

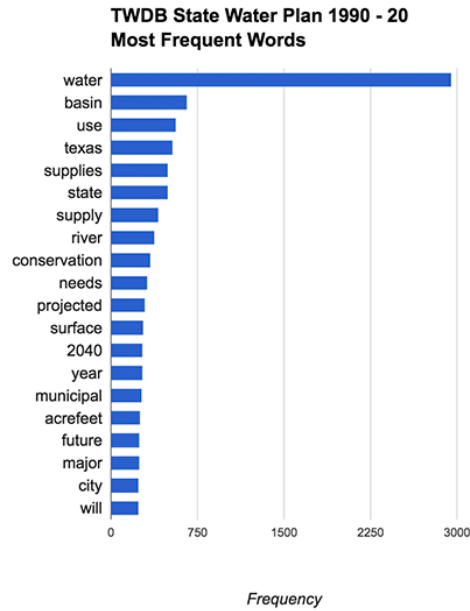
### Word Cloud for 1961 State Water Plan



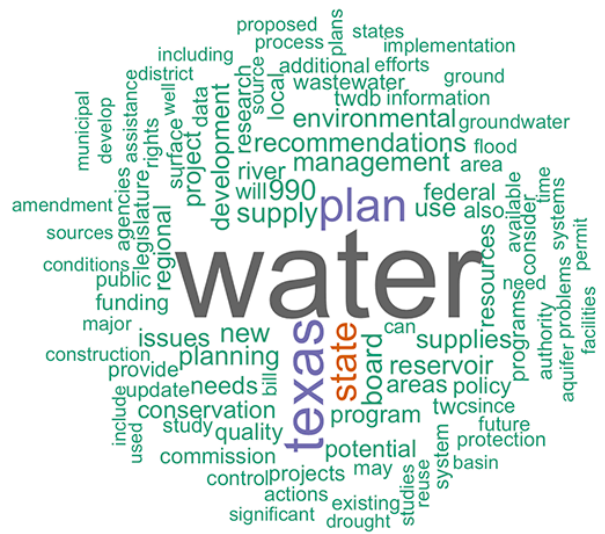
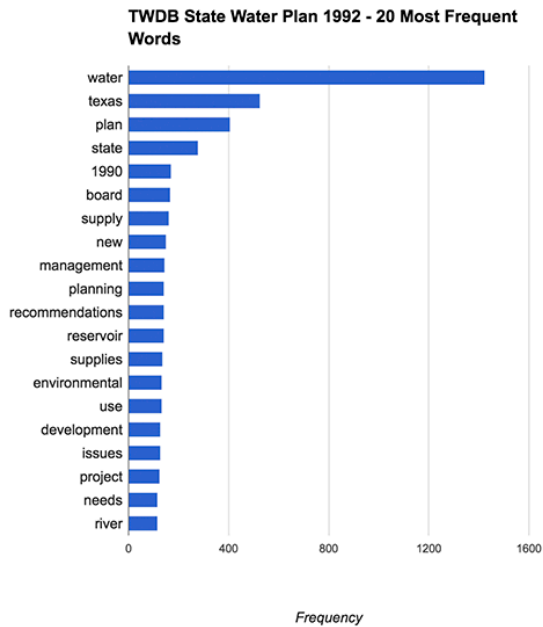
## 5



## Word Cloud for 1990 State Water Plan

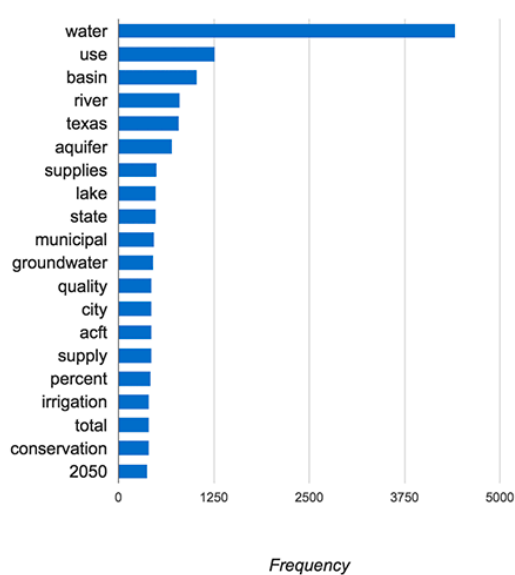


## Word Cloud for 1992 State Water Plan



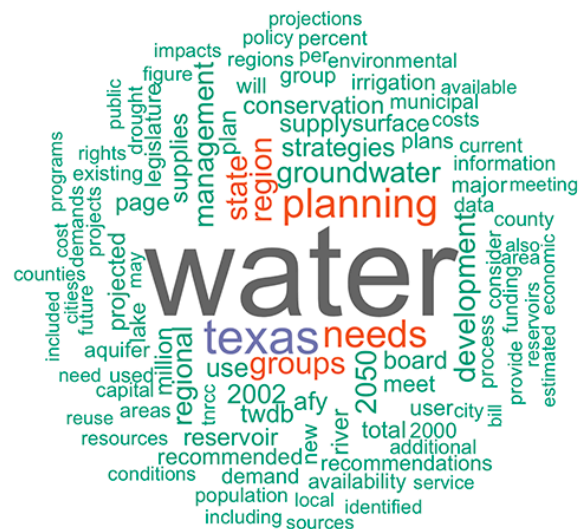
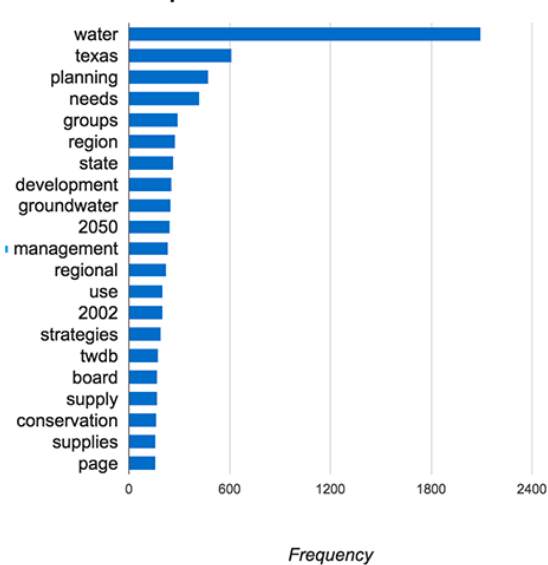
## Word Cloud for 1997 State Water Plan

### TWDB State Water Plan 1997 - 20 Most Frequent Words

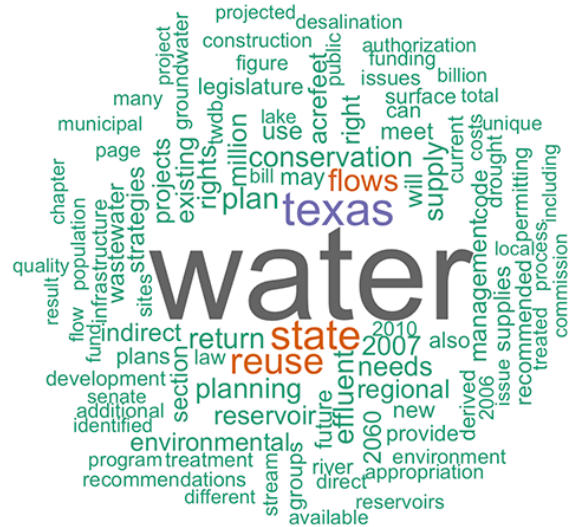
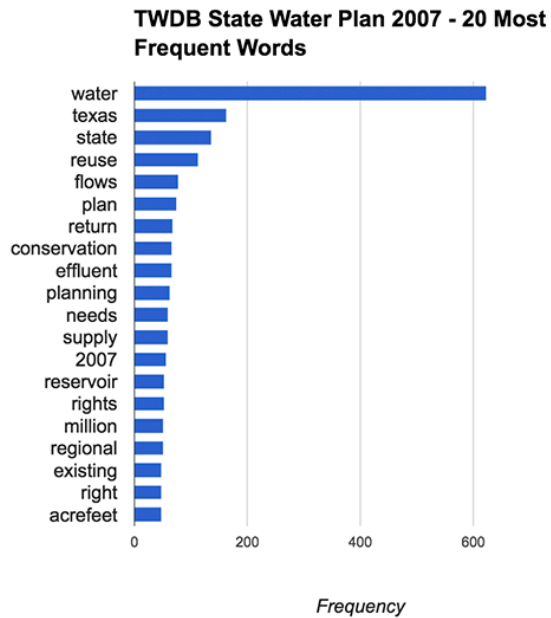


## Word Cloud for 2002 State Water Plan

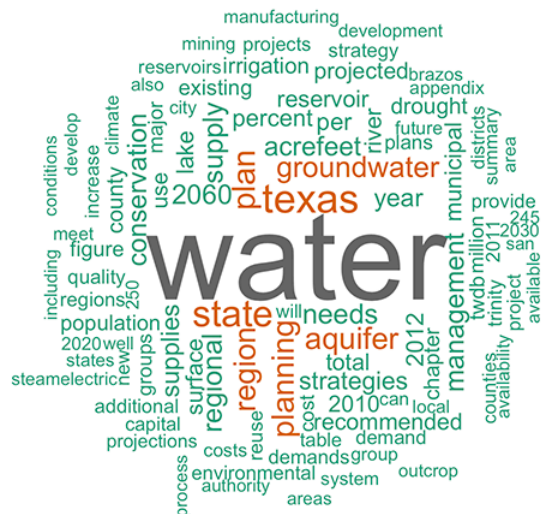
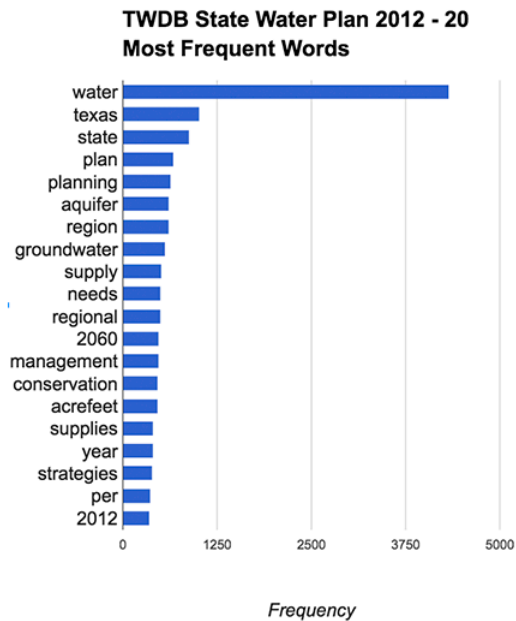
### TWDB State Water Plan 2002 - 20 Most Frequent Words



## Word Cloud for 2007 State Water Plan

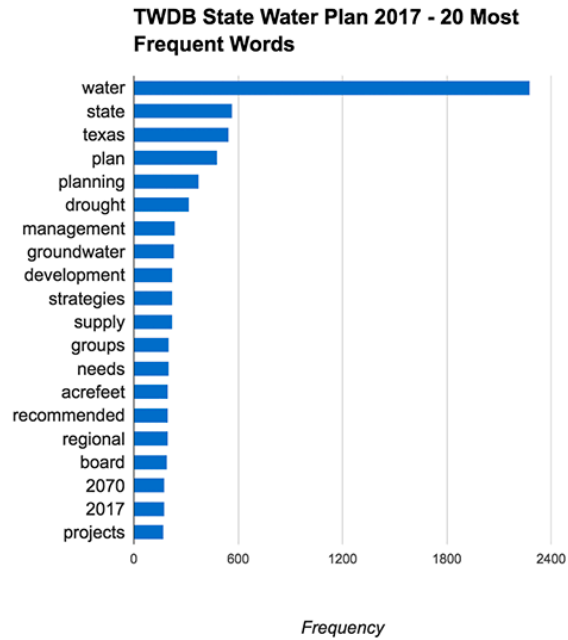


## Word Cloud for 2012 State Water Plan



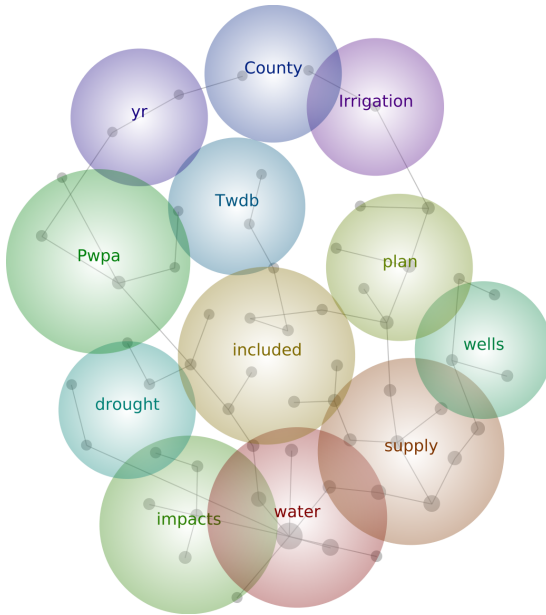


## Word Cloud for 2017 State Water Plan

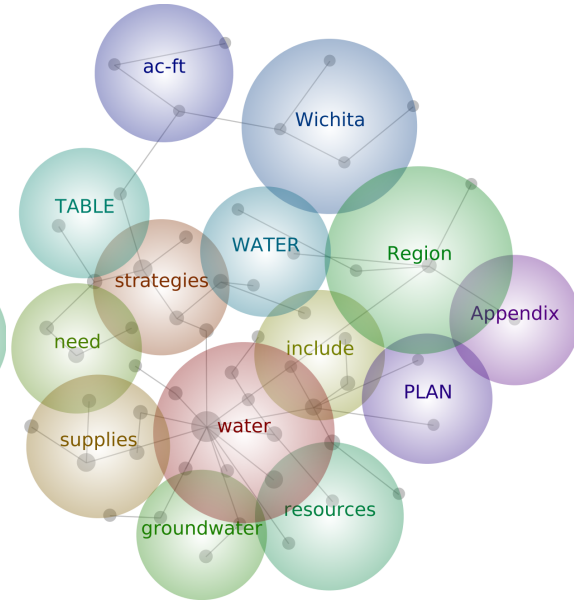


## APPENDIX B - CLUSTER ANALYSIS OF REGIONAL WATER PLANS PROJECTED THROUGH LEXIMANCER

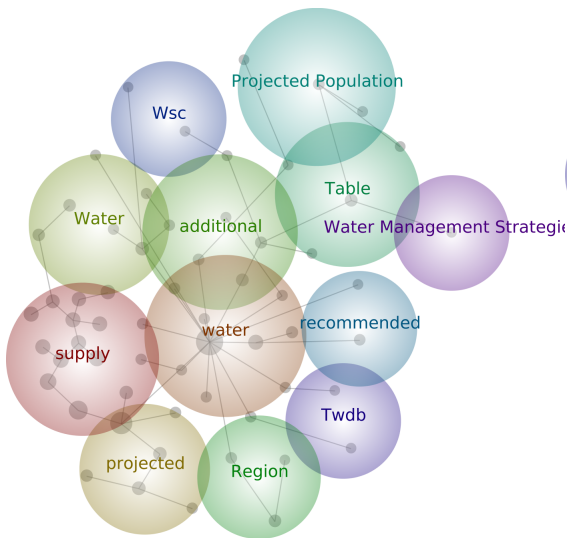
Region A – Panhandle



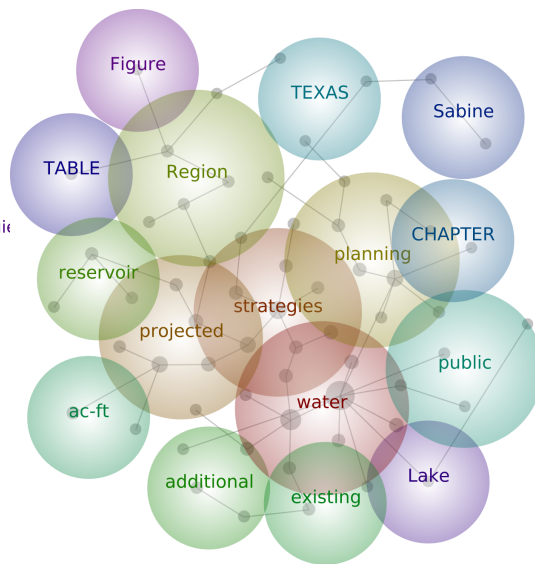
Region B – North Texas



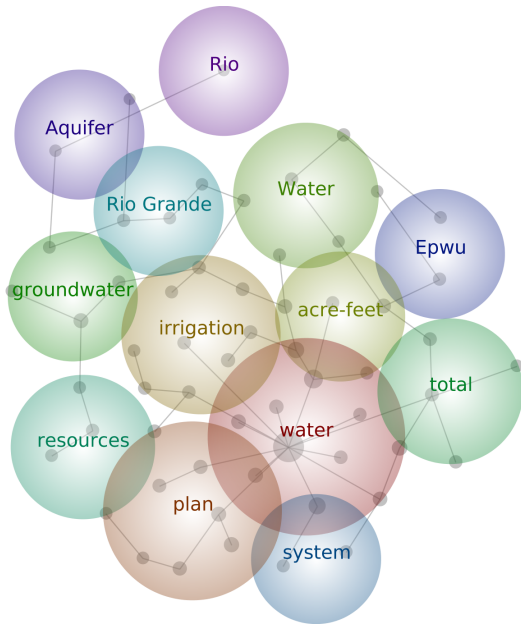
Region C – North Texas



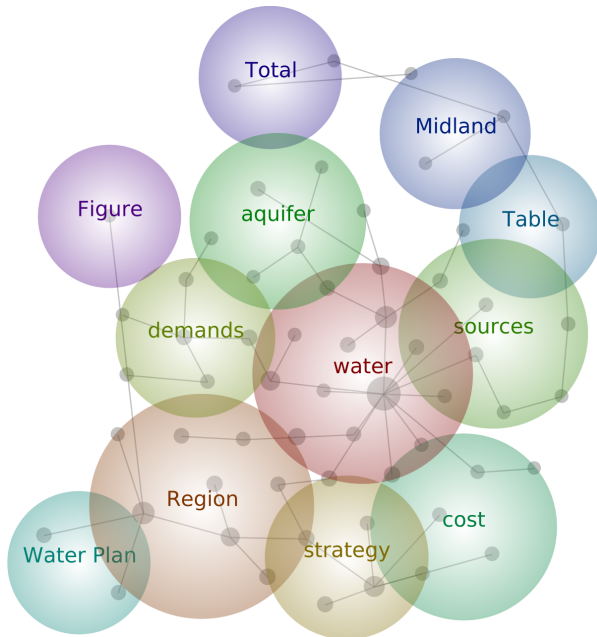
Region D – North East Texas



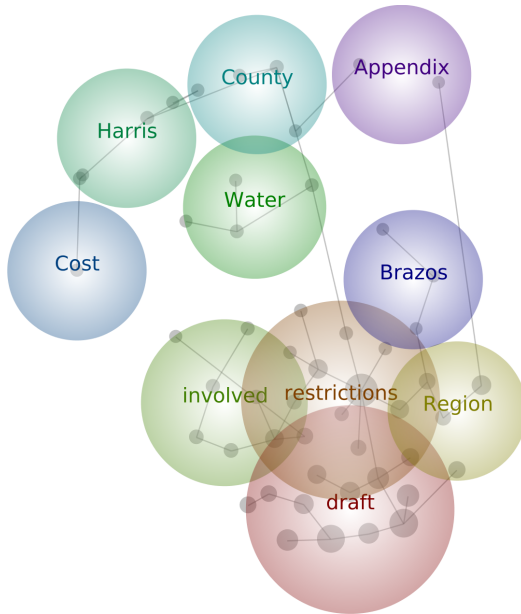
Region E – Far West Texas



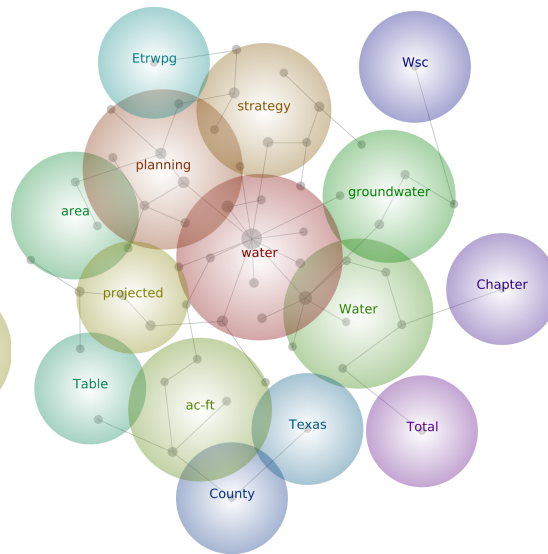
Region F – West Texas



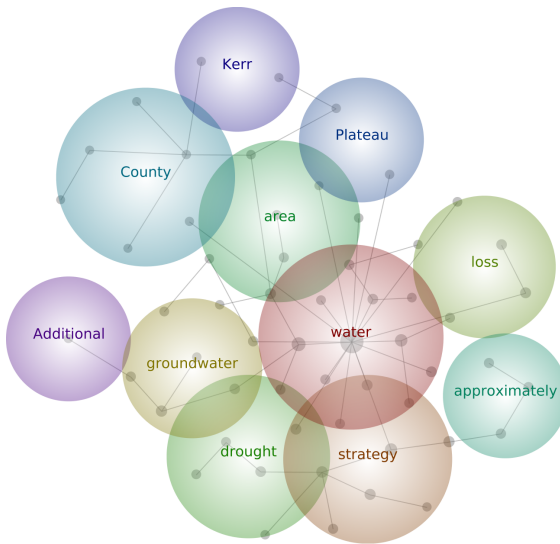
Region H – Coastal



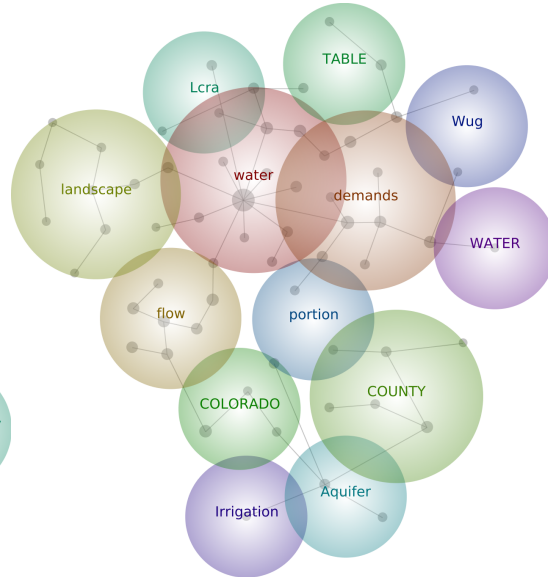
Region I – East Texas



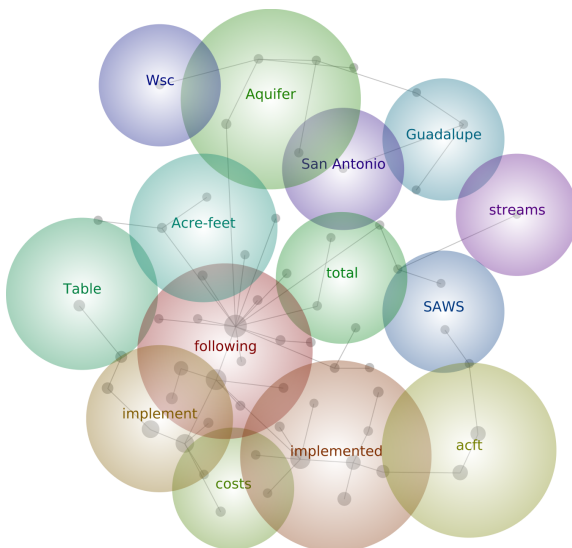
Region J – Plateau



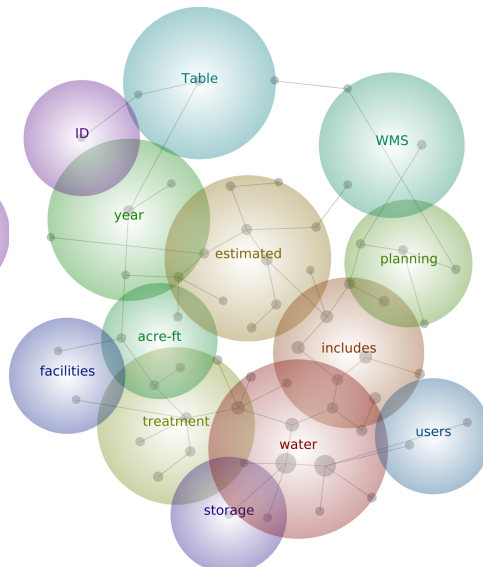
Region K – Lower Colorado



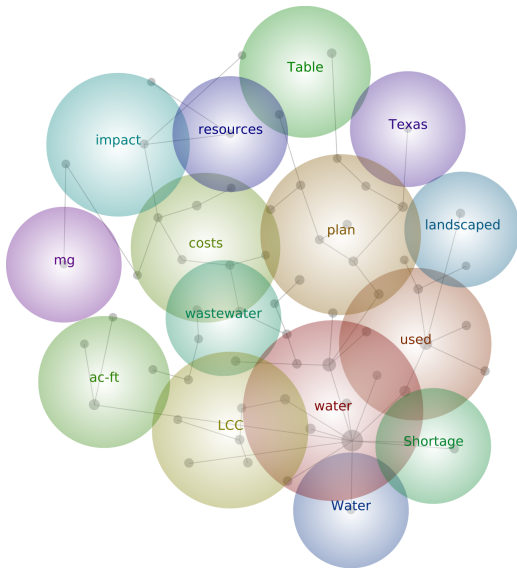
Region L – South Central Texas



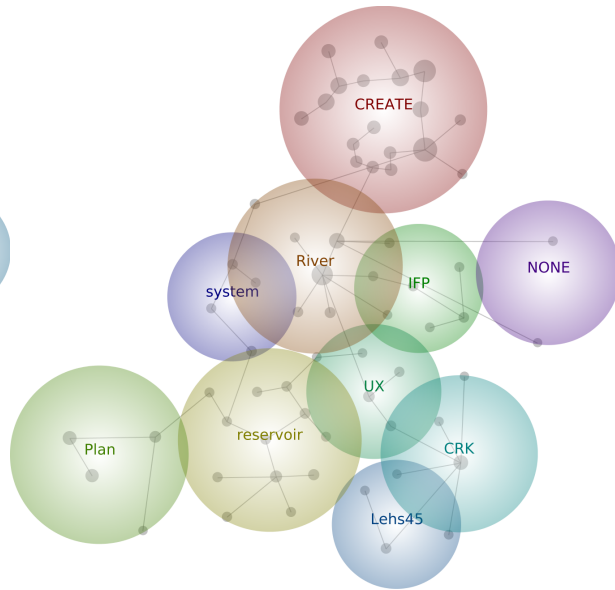
Region M – Rio Grande



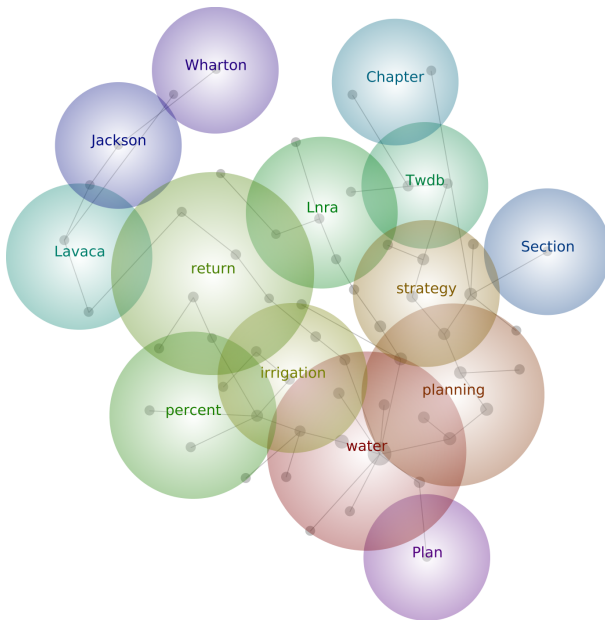
Region N – Coastal Bend



Region O – Llano Estacado



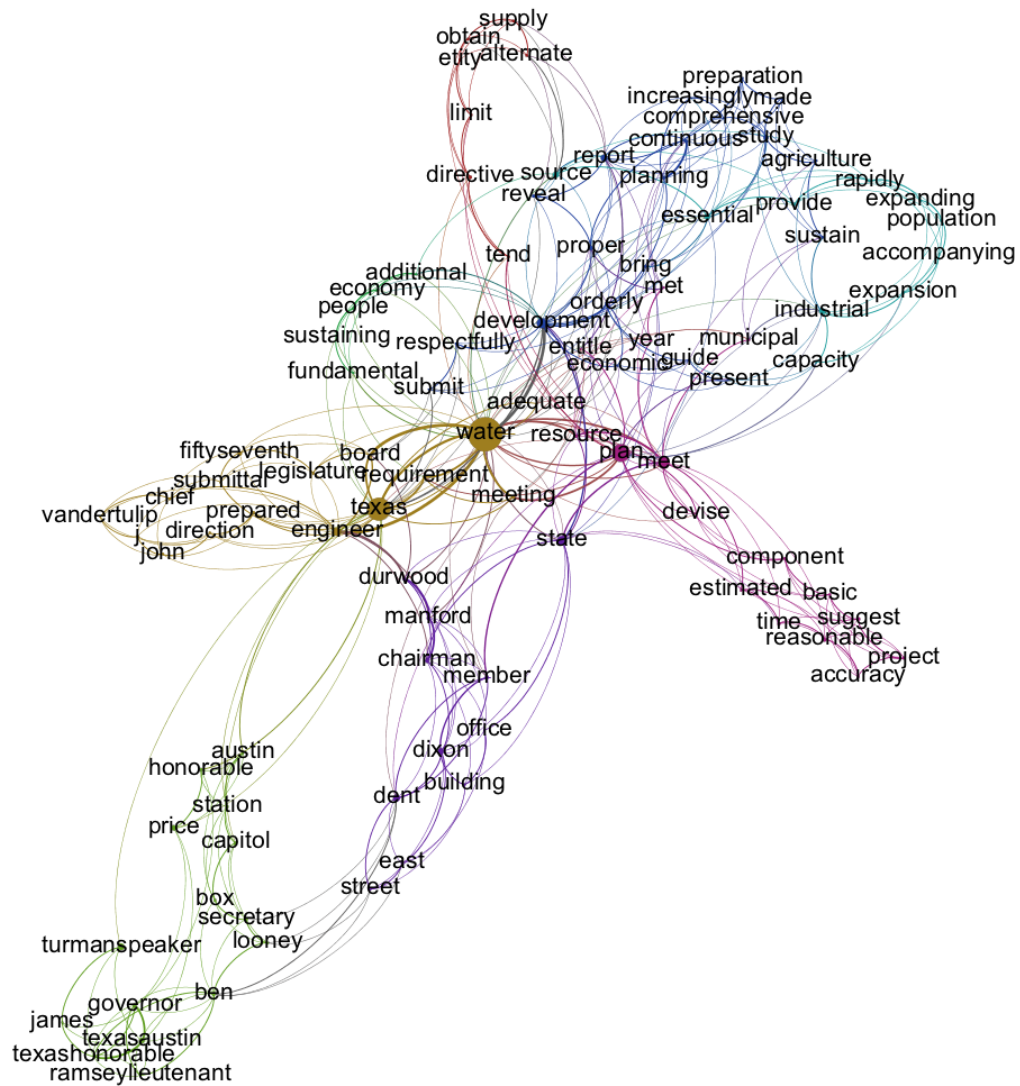
Region P – Lavaca



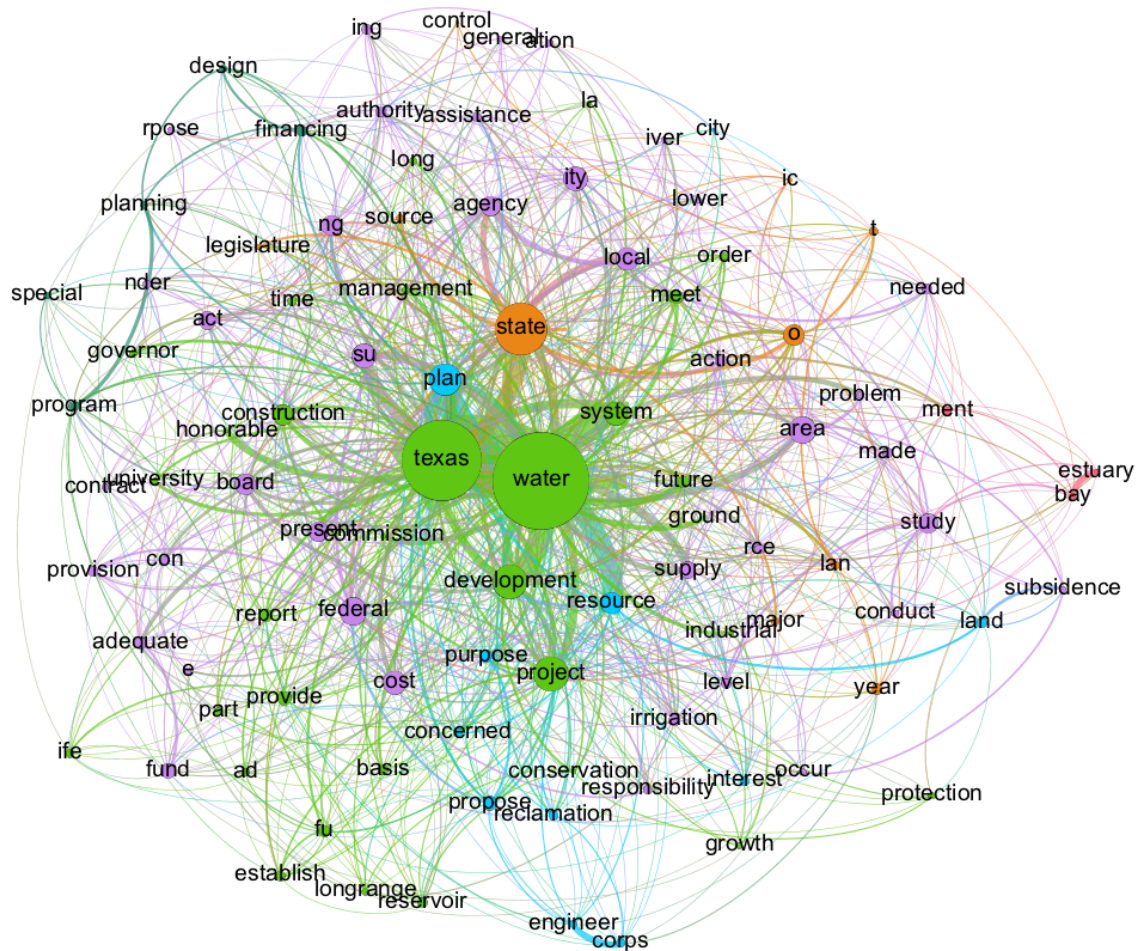


## APPENDIX C - TWDB STATE WATER PLAN NETWORK VISUALIZATIONS USING GEPHI

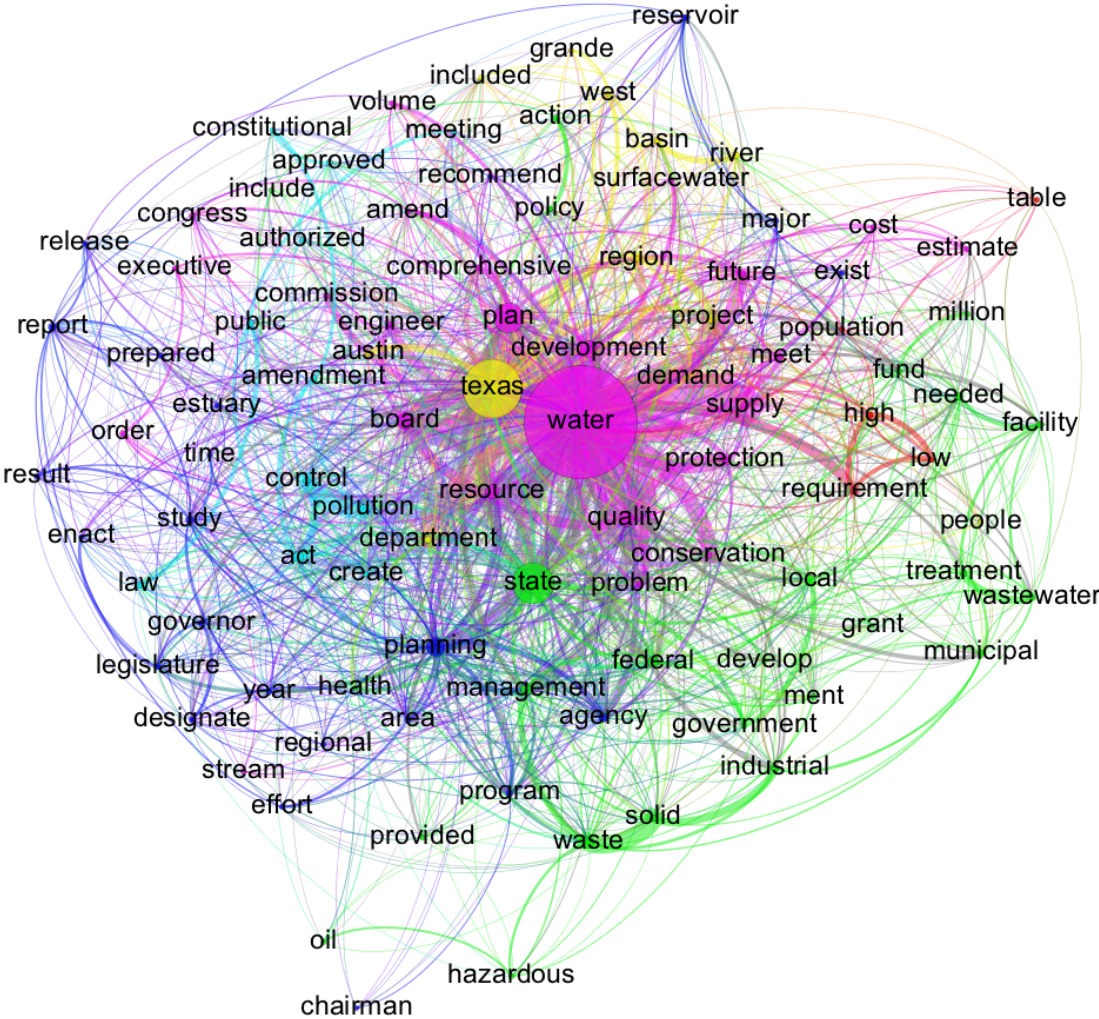
## 1961 State Water Plan



## 1968 State Water Plan

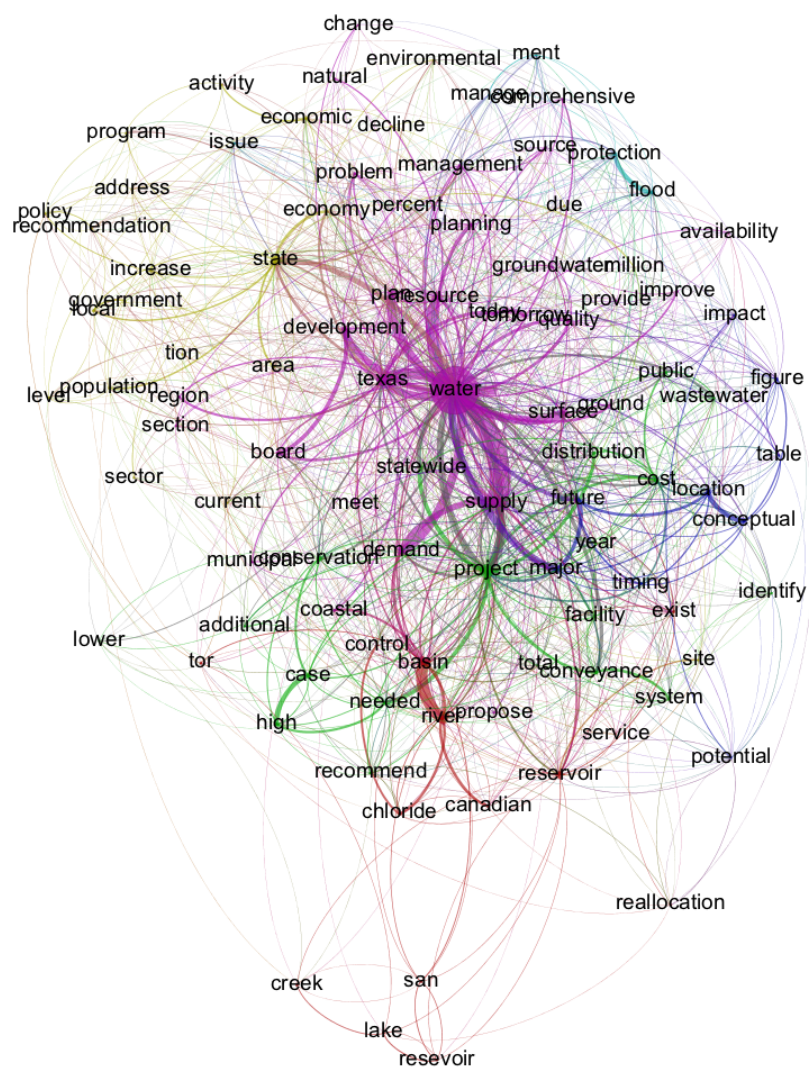


# 1984 State Water Plan

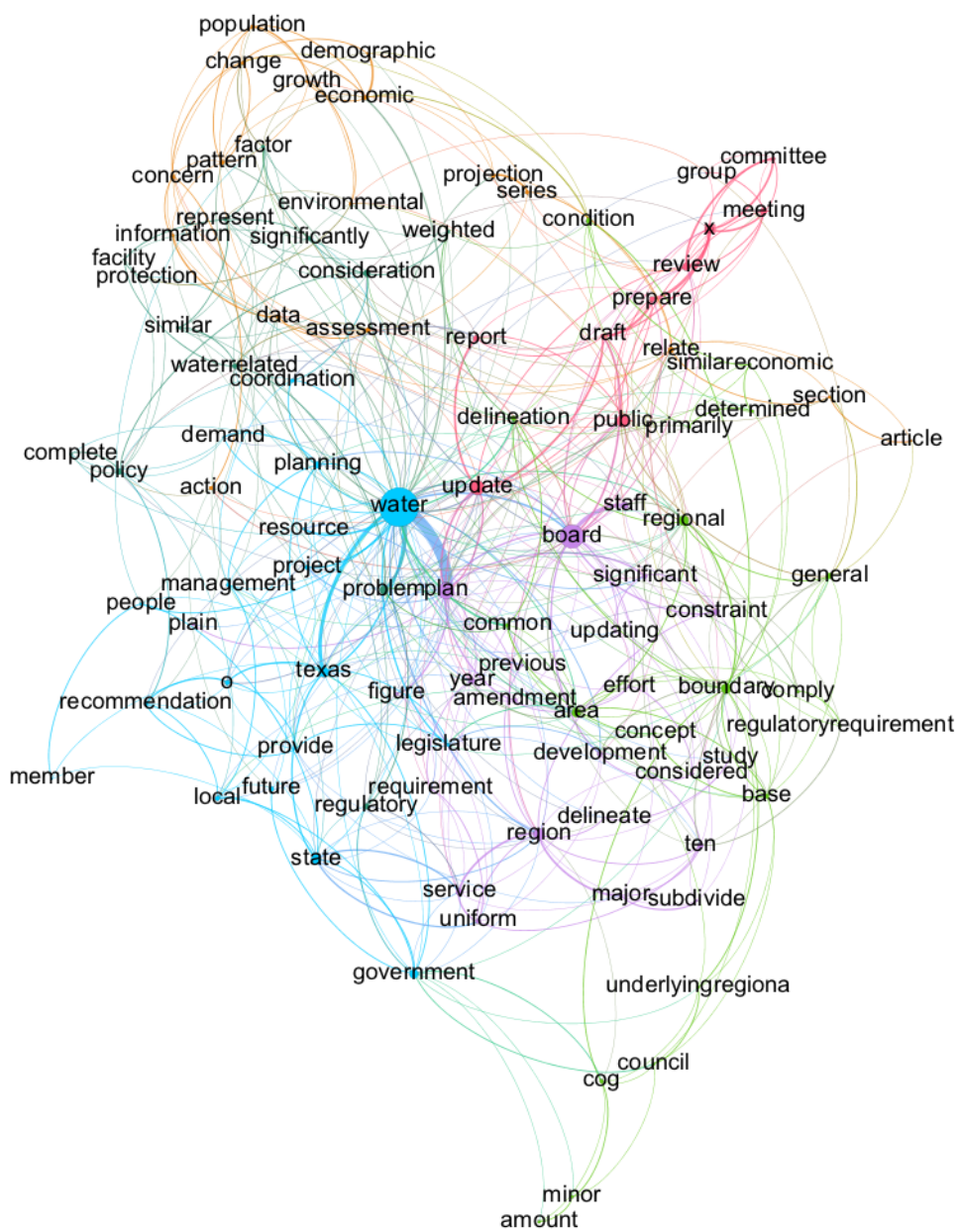




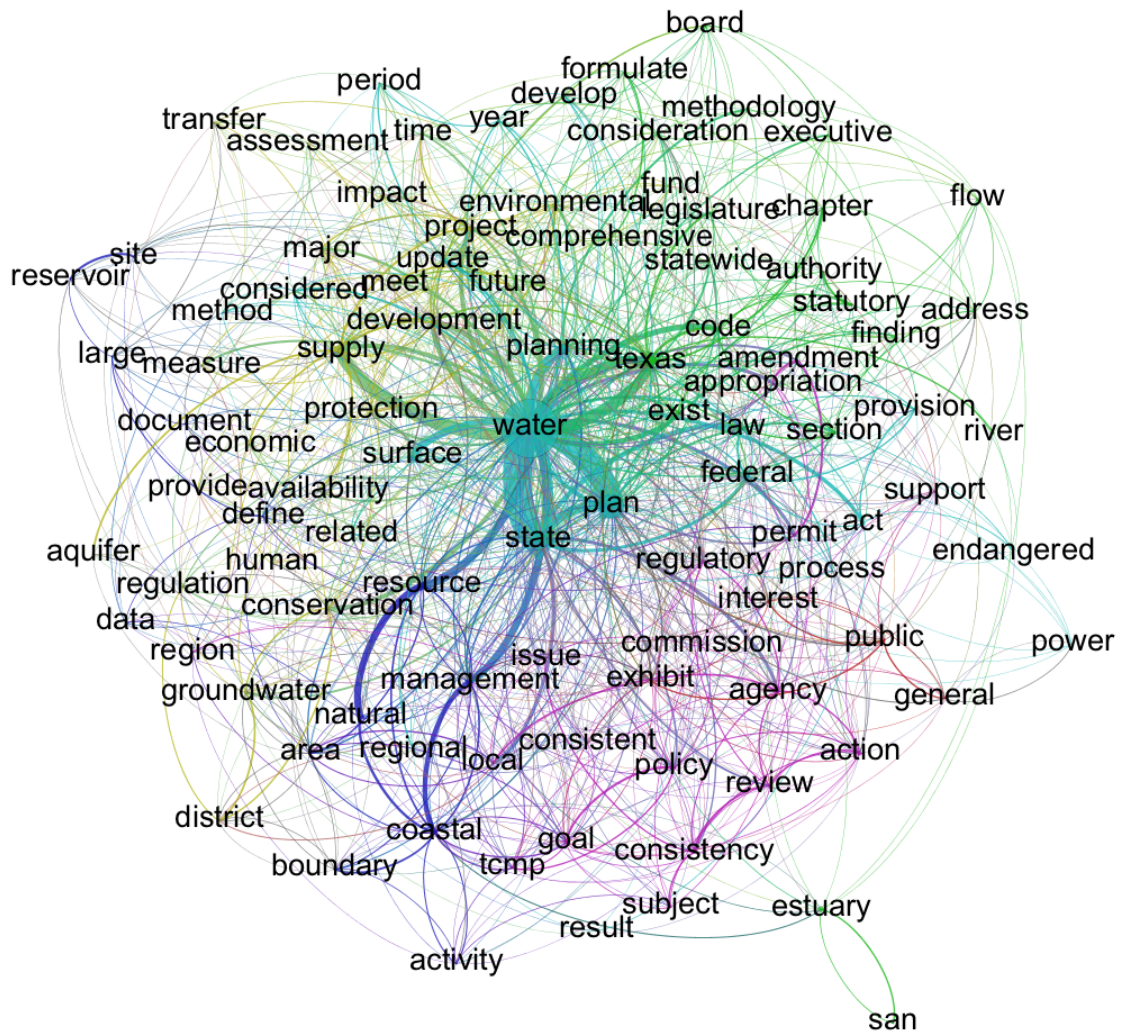
## 1990 State Water Plan



# 1992 State Water Plan

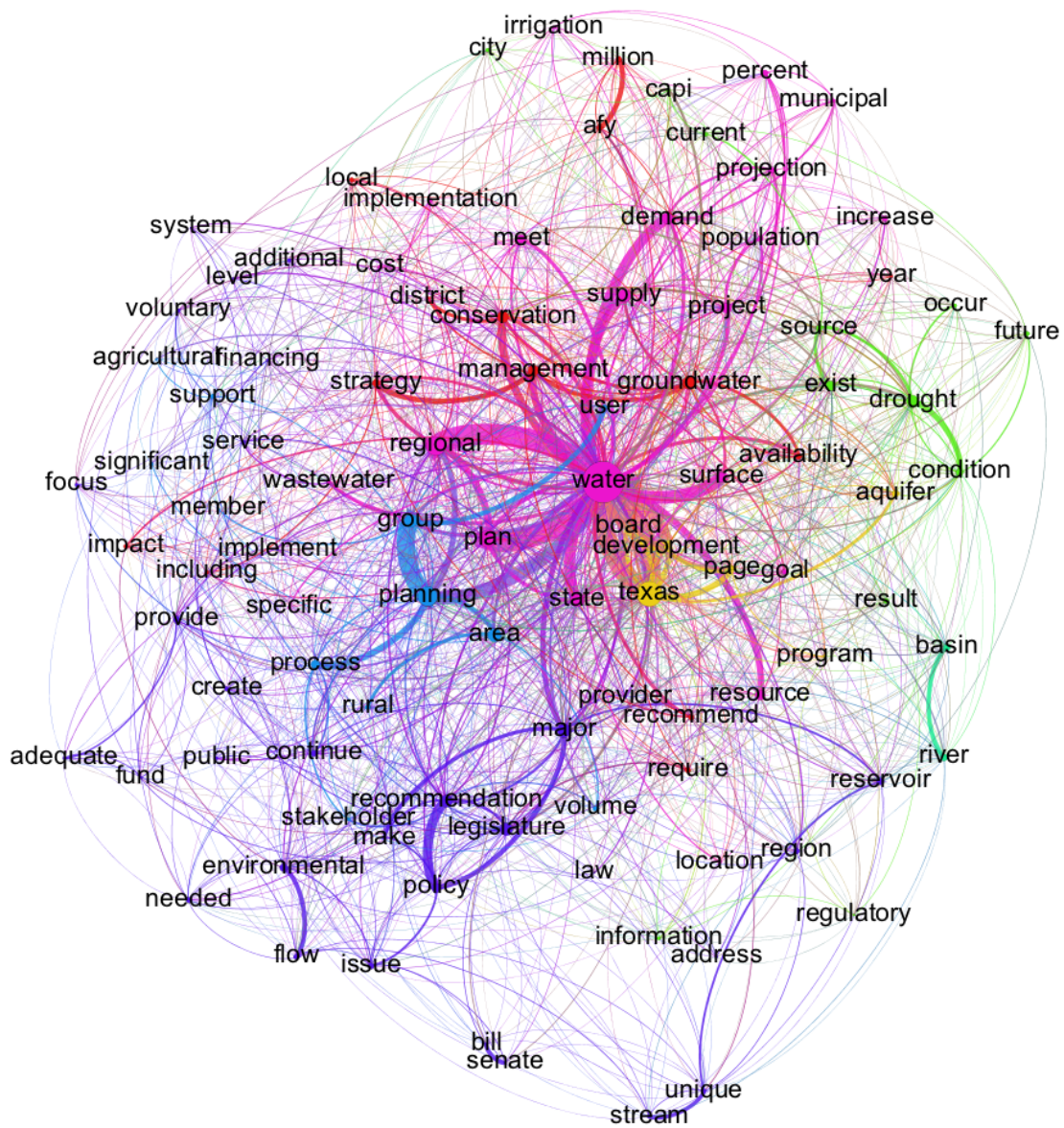


## 1997 State Water Plan

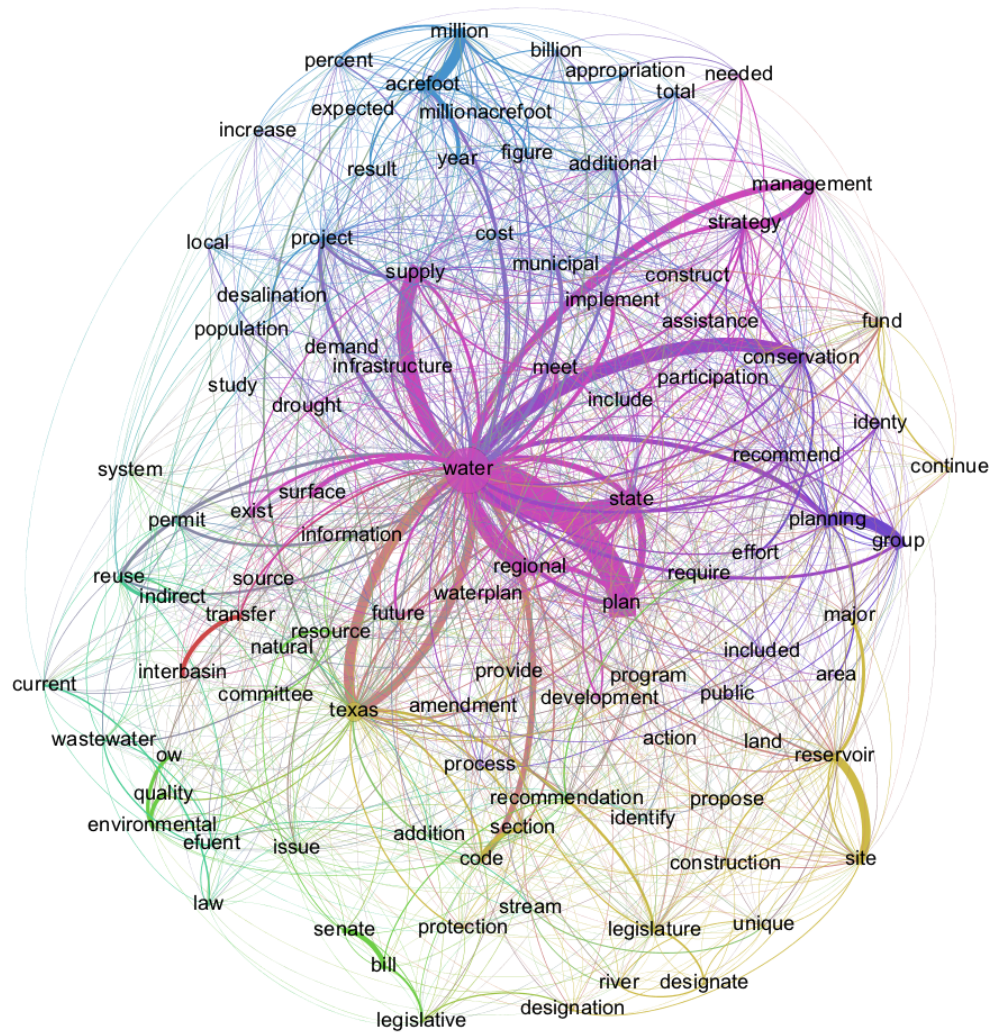




## 2002 State Water Plan

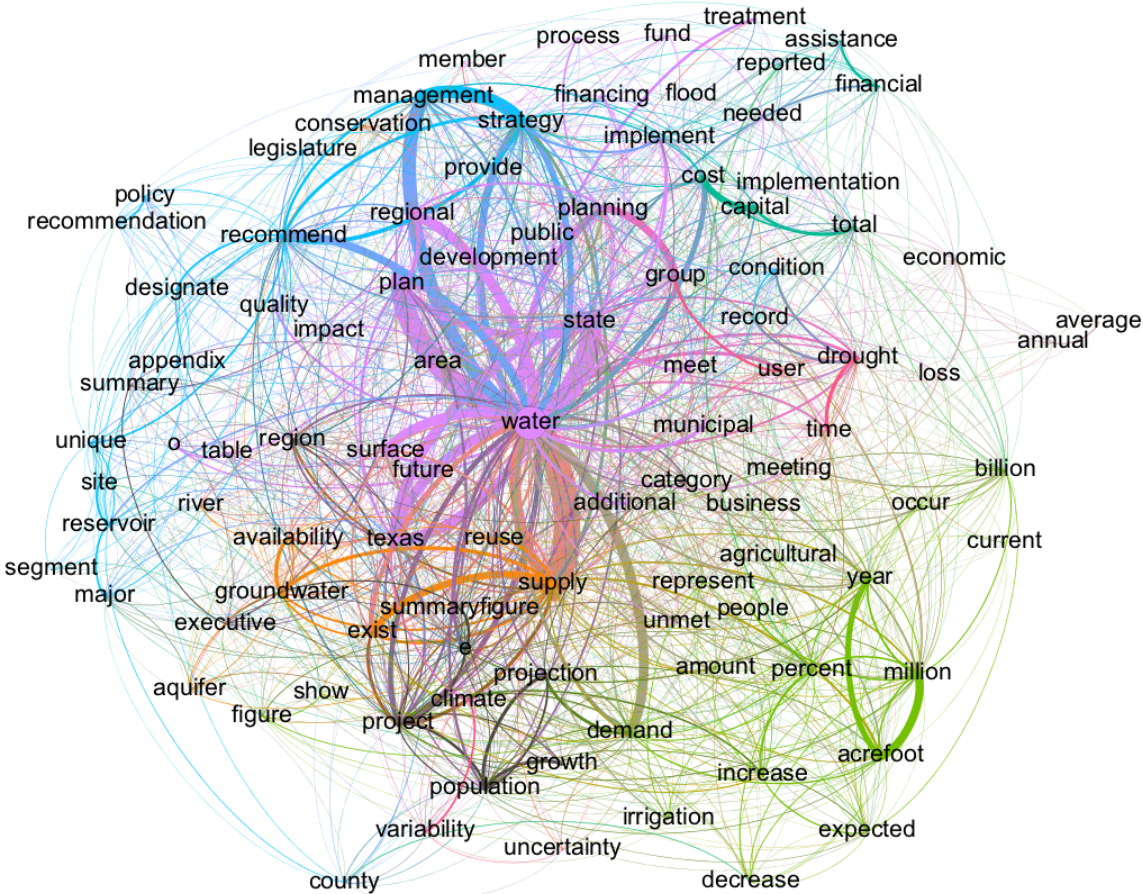


## 2007 State Water Plan

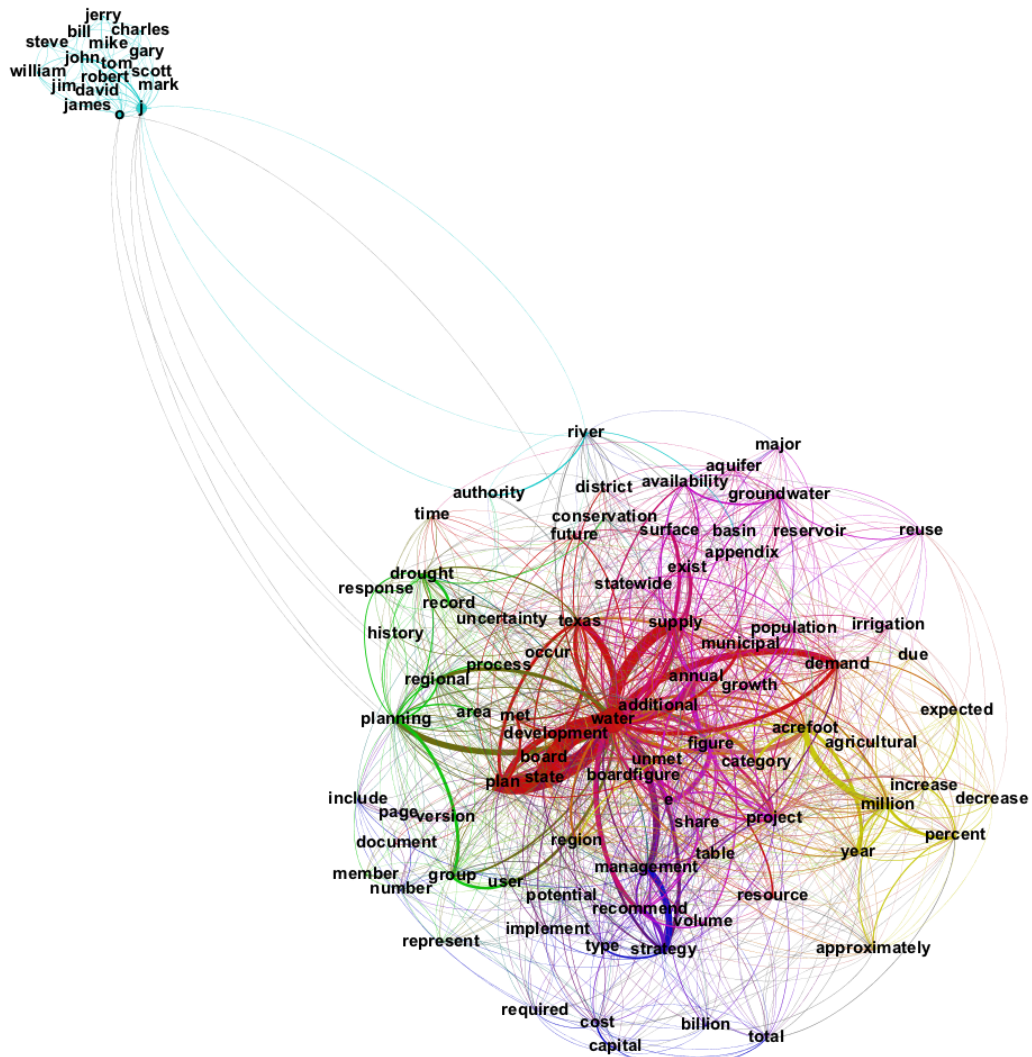




## 2012 State Water Plan



## 2017 State Water Plan



## APPENDIX D - TEXAS GROUNDWATER MANAGEMENT AREAS FUTURE DESIRED CONDITIONS TABLE

Aquifer	Compilation of Desired Future Conditions	Date Desired Future Condition
<b>GMA 1</b>		
Blaine	50 percent of the volume in storage remaining in 50 years in Wheeler County.	6/3/2010
Dockum	Average decline in water levels will decline no more than 30 feet over the next 50 years.	6/3/2010
Ogallala and Rita Blanca	40 percent of volume in storage remaining in 50 years in Dallam, Hartley, Moore, and Sherman counties; 50 percent of volume remaining in 50 years in Armstrong, Potter, Randall, Hansford, Hutchinson, Lipscomb, Ochiltree, Carson, Donley, Gray, Roberts, Wheeler, and Oldham counties; and 80 percent of volume in storage remaining in 50 years in Hemphill County.	7/7/2009
<b>GMA 2</b>		
Dockum	Average water level decline of no more than 40 feet between 2010 and 2060. Not relevant for Dawson, Garza, Howard, Martin, Terry, and Yoakum counties.	8/5/2010
Ogallala and Edwards-Trinity (High Plains)	50 percent of saturated thickness remaining after 50 years for the Northern Portion of Groundwater Management Area 2 (Bailey, Briscoe, Castro, Cochran, Crosby, Deaf Smith, Floyd, Hale, Hockley, Lamb, Lubbock, Lynn, Parmer, and Swisher counties); average water level decline for the Southern Portion of Groundwater Management Area 2 over 50 years by county, Andrews: 6 feet, Bordon: 3 feet, Dawson: 74 feet, Gaines: 70 feet, Garza: 40 feet, Howard: 1 foot, Martin: 8 feet, Terry: 42 feet, and Yoakum: 18 feet.	8/5/2010
<b>GMA 3</b>		
Capitan Reef	Total net decline in water levels over 50 years shall not exceed 200 feet below water levels in the aquifer in the year 2010. Not relevant in Crane and Loving counties.	8/9/2010
Dockum	Average total net decline in water levels over 50 years shall not exceed 27 feet below water levels in the aquifer in the year 2010.	8/9/2010
Edwards-Trinity (Plateau) and Pecos Valley	Average total net decline in water levels over 50 years shall not exceed 28 feet below water levels in the aquifers in 2010.	8/9/2010



Rustler	Average total net decline in water levels within the unconfined portion in Reeves County over 50 years shall not exceed 15 feet below water levels in the aquifer in 2010; and the average total net decline in water levels within the confined portion in Pecos, Loving, Reeves and Ward counties over 50 years shall not exceed 300 feet below water levels in the aquifer in the year 2010. Not relevant in Crane and Winkler counties.	8/9/2010
<b>GMA 4</b>		
Bone Spring-Victorio Peak	Hudspeth County Underground Water Conservation District No 1: 0 foot drawdown.	8/13/2010
Capitan Reef	Brewster County Groundwater Conservation District: 0 foot drawdown, Culberson County Groundwater Conservation District: 50 feet of drawdown. Not relevant in Jeff Davis and Hudspeth counties.	8/13/2010
Edwards-Trinity (Plateau)	Brewster County GCD: 3 feet of drawdown, Culberson County GCD: 50 feet of drawdown. Not relevant in Jeff Davis County.	8/13/2010, amended 5/19/11
Igneous	Brewster County Groundwater Conservation District: 10 feet of drawdown, Culberson County Groundwater Conservation District: 66 feet of drawdown, Jeff Davis County Groundwater Conservation District: 20 feet of drawdown, Presidio County Groundwater Conservation District: 14 feet of drawdown.	8/13/2010
Marathon	Brewster County Groundwater Conservation District: 0 foot drawdown. Not relevant in Culberson and Jeff Davis counties.	8/13/2010
Presidio-Redford Bolson	Presidio County Groundwater Conservation District: 5 feet of drawdown.	8/13/2010
Rustler	Brewster County Groundwater Conservation District: 0 foot drawdown. Not relevant in Culberson and Jeff Davis counties.	8/13/2010
West Texas Bolsons	Culberson County Groundwater Conservation District: 78 feet of drawdown, Jeff Davis County Groundwater Conservation District: 72 feet of drawdown, Presidio County Groundwater Conservation District: 72 feet of drawdown. Not relevant in Hudspeth County.	8/13/2010
Upper Salt Basin	Culberson County Groundwater Conservation District: 50 feet of drawdown.	8/13/2010
<b>GMA 5</b>	No future groundwater desired conditions	
<b>GMA 6</b>		
Blaine	Clear Fork Groundwater Conservation District (Fisher County): total decline in water levels will be no more than 4 feet over the next 50 years; Gateway Groundwater Conservation District (Childress, Cottle, Foard, and Hardeman counties): total decline in water levels will be no more than 2 feet over the next 50 years; Mesquite Groundwater Conservation District (Childress, Collingsworth and Hall counties): 80 percent of current volume of storage remaining in 50 years; King County: total decline in water levels will be no more than 7 feet over the next 50 years. Not relevant in Dickens, Knox, Motley, Stonewall, and Wilbarger counties.	7/22/10, amended 7/19/2011

Dockum	Clear Fork Groundwater Conservation District (Fisher County): total decline in water levels will be no more than 25 feet over the next 50 years; Gateway Groundwater Conservation District (Motley County), Dickens and Kent counties: total decline in water levels will be no more than 40 feet over the next 50 years.	7/22/2010
Ogallala	Motley (Gateway Groundwater Conservation District) and Dickens counties: 50 percent of volume in storage remaining in 50 years.	7/22/2010
Seymour	"Pods 1, 2, and 3" in Mesquite Groundwater Conservation District (Collingsworth, Childress, and Hall counties): 50 percent of current volume in storage remaining in 50 years. "Pods 3 and 4" in Gateway Groundwater Conservation District (Motley, Childress, Foard, Hardeman counties): total decline in water levels will be no more than 1 foot over 50 years; "Pod 4" in Wichita and Wilbarger counties: total decline in water levels will be no more than 1 foot over 50 years; "Pod 5" in Archer, Clay, Wichita, and Wilbarger counties: total decline in water levels will be no more than 2 feet over 50 years; "Pods 6, 7, and 8" in Rolling Plains Groundwater Conservation District (Baylor, Knox, and Haskell counties): total decline in water levels will be no more than 18 feet over 50 years; "Pod 7" in Stonewall County: total decline in water levels will be no more than 24 feet over 50 years; "Pod 8" in Throckmorton and Young counties: total decline in water levels will be no more than 3 feet over 50 years; "Pods 9 and 10" in Kent and Stonewall counties: total decline in water levels will be no more than 4 feet over 50 years; "Pod 11" in Clear Fork Groundwater Conservation District (Fisher County): total decline in water levels will be no more than 1 foot over 50 years; "Pods 11 through 15" in Jones and Stonewall counties: total decline in water levels will be no more than 1 foot over 50 years. "Pod 1" in Gateway Groundwater Conservation District (Childress County) is not relevant.	7/22/10, amended 7/19/2011
<b>GMA 7</b>		
Capitan Reef	Total net decline in water levels within the Middle Pecos Groundwater Conservation District over 50 years shall not exceed 15 below water levels in the unconfined portion of the aquifer in the year 2010; and total net decline in water levels over 50 years shall not exceed 200 feet below water levels in the confined portion in the aquifer in year 2010. Not relevant outside of district boundaries.	7/29/2010
Dockum	Upper Dockum: net total drawdown not to exceed 29 feet in Midland County. Lower Dockum: net total drawdown not to exceed 4 feet in Ector, Mitchell, Pecos, Scurry, and Upton counties (Lone Wolf Groundwater Conservation District, Middle Pecos Groundwater Conservation District); and drawdown not to exceed a net total of 39 feet in Nolan County (West-Tex Groundwater Conservation District). Not relevant in all other areas of Groundwater Management Area 7.	7/29/2010
Edwards-Trinity (Plateau), [Trinity, and Pecos Valley]	Average drawdown of 7 feet except within Kinney County GCD. Kinney County drawdown consistent with maintaining annual average flow of 23.9 cubic feet per second and median flow of 24.4 cubic feet per second at Los Moras Springs.	7/29/2010
Ellenburger-San Saba	Total net decline in water levels within Hickory Underground Water Conservation District No. 1, Hill Country Underground Water Conservation District, Kimble County Groundwater Conservation District, and Menard County Underground Water District over 50 years shall not exceed 5 feet below 2010 levels. Not relevant in all other areas of Groundwater Management Area 7.	7/29/2010

Hickory	Total net decline in water levels within Hickory Underground Water Conservation District No. 1, Hill Country Underground Water Conservation District, Kimble County Groundwater Conservation District, Menard County Underground Water District, and Llano County and non-district areas in McCulloch and San Saba counties over 50 years shall not exceed 7 feet below 2010 levels. Not relevant in all other areas of Groundwater Management Area 7.	7/29/2010
Lipan	Within Lipan-Kickapoo Water Conservation District in Concho, Runnels, and Tom Green counties continue to use 100 percent of all available groundwater annually with annual fluctuations of water levels and zero net drawdown in water levels over the next 50 years. Not relevant outside of district boundaries.	7/29/2010
Marble Falls	Total net decline in water levels in San Saba County over 50 years shall not exceed 7 feet below 2010 water levels in the aquifer. Not relevant in all other areas of Groundwater Management Area 7.	7/29/2010
Ogallala	Total decline in volume of water within Ector, Glasscock, and Midland counties over 50 years shall not exceed 50 percent of volume in the aquifer in the year 2010. Not relevant in all other areas of Groundwater Management Area 7.	7/29/2010
Rustler	Total net decline in water levels within the Middle Pecos Groundwater Conservation District over 50 years shall not exceed 300 feet below water levels in the aquifer in year 2010. Not relevant outside of district boundaries.	7/29/2010
<b>GMA 8</b>		
Blossom	From estimated year 2009 conditions, Bowie County: average drawdown of the unconfined zone should not exceed approximately 5.4 feet after 50 years; Lamar County: average drawdown of the unconfined zone should not exceed approximately 2.4 feet after 50 years; Red River County: average drawdown of the unconfined zone should not exceed approximately 6.5 feet after 50 years; Bowie, Lamar, and Red River counties: drawdown of the confined zone should not exceed approximately 20 feet after 50 years	4/27/2011
Brazos River Alluvium	Maintain approximately 100 percent of the saturated thickness after 50 years in Falls County; maintain approximately 82 percent of estimated saturated thickness after 50 years in McLennan County; and maintain approximately 90 percent of the estimated saturated thickness after 50 years in Hill and Bosque counties. Not relevant in Milan County.	4/27/2011, amended 6/23/2011
Edwards (BFZ)	Maintain at least 100 acre-feet per month of stream/spring flow in Salado Creek during a repeat of the drought of record in Bell County; Maintain at least 42 acre-feet per month of aggregated stream/spring flow during a repeat of the drought of record in Travis County; and Maintain at least 60 acre-feet per month of aggregated stream/spring flow during a repeat of the drought of record in Williamson County.	4/27/2011
Ellenburger-San Saba	Burnet County: maintain approximately 100 percent of the saturated thickness after 50 years by using approximately 80 percent of the estimated recharge; Lampasas County: maintain approximately 90 percent of the saturated thickness after 50 years; Brown and Mills counties: maintain approximately 90 percent of the available drawdown after 50 years.	4/27/2011

Hickory	Burnet County: maintain approximately 100 percent of the saturated thickness after 50 years by using approximately 80 percent of the estimated recharge; Brown, Lampasas, Mills, Travis, and Williamson counties: maintain approximately 90 percent of the available drawdown [saturated thickness] after 50 years.	4/27/2011
Marble Falls	Burnet County: maintain approximately 100 percent of the saturated thickness after 50 years by using approximately 80 percent of the estimated recharge; Lampasas County: maintain approximately 90 percent of the saturated thickness after 50 years.	4/27/2011
Nacatoch	Drawdown by county: Bowie County: 10 feet in the Red River Basin, 17 feet in the Sulphur River Basin; Delta County: 5 feet; Ellis County: 4 feet; Franklin County: 6 feet; Hopkins County: 10 feet in the Sabine River Basin, 12 feet in the Sulphur River Basin; Hunt County: 10 feet in the Sabine River Basin, 6 feet in the Sulphur River Basin; Kaufman County: 7 feet in the Sabine River Basin, 4 feet in the Trinity River Basin; Lamar County: 5 feet; Navarro County: 4 feet; Rains County: 13 feet; Red River County: 10 feet in the Red River Basin, 8 feet in the Sulphur River Basin; and Rockwall County: 5 feet.	6/23/2011
Trinity	Listed DFCs by county and aquifer layers (Paluxy, Glen Rose, Hensell, Hosston): From estimated year 2000 conditions, the average drawdown after 50 years should not exceed approximately: see table to the right.	4/27/2011
Woodbine	From estimated year 2000 conditions, the average drawdown after 50 years should not exceed approximately: Colin County: 154 feet, Cooke County: 0 feet, Dallas County: 112 feet, Denton County: 16 feet, Ellis County: 102 feet, Fannin County: 186 feet, Grayson County: 28 feet, Hill County: 87 feet, Hunt County: 353 feet, Johnson County: 4 feet, Kaufman County: 211 feet, Lamar County: 297 feet, Navarro County: 177 feet, Red River County: 202 feet, Rockwall County: 241 feet, Tarrant County: 2 feet. Non-relevant in McLennan County.	4/27/2011, amended 6/23/2011
<b>GMA 9</b>		
Edwards Group of Edwards-Trinity (Plateau)	No net increase in average drawdown in Kendall and Bandera counties. Not relevant in Kerr and Blanco counties.	7/26/2010
Ellenburger-San Saba	Allow for an increase in average drawdown of no more than 2 feet [in Blanco County].	8/29/2008
Hickory	Allow for an increase in average drawdown of no more than 7 feet [in Blanco County].	8/29/2008
Marble Falls	Allow for no net increase in average drawdown [in Blanco County].	8/29/2008
Trinity	Allow for an increase in average drawdown of approximately 30 feet through 2060.	7/26/2010
<b>GMA 10</b>		
Austin Chalk (Uvalde County)	No drawdown (including exempt and non-exempt use).	8/23/2010

Buda Limestone (Uvalde County)	No drawdown (including exempt and non-exempt use).	8/23/2010
Edwards (BFZ) Northern Subdivision	Springflow at Barton Springs during average recharge conditions shall be no less than 49.7 cubic feet per second averaged over an 84 month (7-year) period; and during extreme drought conditions, including those as severe as a recurrence of the 1950s drought of record, springflow of Barton Springs shall be no less than 6.5 cubic feet per second averaged on a monthly basis.	8/4/2010
Edwards (BFZ) Northern Subdivision Saline Zone	Well drawdown at the saline-freshwater interface (the so-called Edwards Bad Water Line) averages no more than 5 feet and does not exceed a maximum of 25 feet at any one point on the interface.	8/4/2010
Edwards (BFZ) San Antonio Segment within Edwards Aquifer Authority	Desired future conditions and modeled available groundwater for the Edwards Aquifer within jurisdiction of the Edwards Aquifer Authority are set by the Texas Legislature (Act of May 28, 2007, 80th Leg., R.S., ch. 1351, § § 2.02 and 2.06, 2007 Tex. Gen. Laws, 4612, 4627, and 4627; Act of May 28, 2007, 80th Leg., R.S. ch. 1430, § § 12.02 and 12.06, 2007 Tex. Gen. Laws 5848, 5901, and 5903). The DFCs are specified in Sections 1.14(a), (f), (h), and 1.26 of the Edwards Aquifer Authority Act. The DFCs are specified in Sections 1.14(a), (f), (h), and 1.26 of the Edwards Aquifer Authority Act, and relate to levels in index wells (J-17 in the San Antonio pool and J-27 in the Uvalde pool) or flows in the Comal Springs and San Marcos Springs. Refer to the Edwards Aquifer Authority Groundwater Management Plan for details.	5/28/2007
Edwards (Kinney County)	Water level in well number 70-38-902 shall not fall below 1184 feet mean sea level.	8/4/2010
Leona Gravel (Medina County)	Average drawdown of 15 feet.	5/17/2010
Leona Gravel (Uvalde County)	No drawdown (including exempt and non-exempt use).	8/23/2010
Trinity	Average regional well drawdown not exceeding 25 feet during average recharge conditions (including exempt and non-exempt use); within Hays-Trinity Groundwater Conservation District: no drawdown; within Uvalde County: 20 feet. Not relevant in Trinity-Glen Rose GCD. Note: Hays-Trinity Groundwater Conservation District and Trinity-Glen Rose Groundwater Conservation District are no longer within the Groundwater Management Area 10 boundary.	8/23/2010
<b>GMA 11</b>		
Yegua Jackson, Sparta, Weches, Queen City, Reklaw and Carrizo-Wilcox	Allowing up to an average drawdown of 17 feet.	4/13/2010
<b>GMA 12</b>		

Brazos River Alluvium	Milan County: a decrease of 5 feet in average saturated thickness from 2010 to 2060. The baseline thickness for 2010 is estimated at 24.5 feet; Burleson County: a decrease of 6 feet in average saturated thickness from 2010 to 2060. The baseline thickness for 2010 is estimated at 38.5 feet. Not relevant in Brazos Valley GCD.	8/11/2010
Calvert Bluff (Upper Wilcox)	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 106 feet; Lost Pines Groundwater Conservation District: 99 feet; Mid-East Texas Groundwater Conservation District: 70 feet; Post Oak Savannah Groundwater Conservation District: 140 feet; Limestone County: 9 feet, Navarro County: 0 feet, Williamson County: 10 foot water level rise.	8/11/2010
Carrizo	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 47 feet; Fayette County Groundwater Conservation District: 60 feet; Lost Pines Groundwater Conservation District: 47 feet; Mid-East Texas Groundwater Conservation District: 55 feet; Post Oak Savannah Groundwater Conservation District: 65 feet.	8/11/2010
Hooper (Lower Wilcox)	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 170 feet; Lost Pines Groundwater Conservation District: 129 feet; Mid-East Texas Groundwater Conservation District: 95 feet; Post Oak Savannah Groundwater Conservation District: 180 feet; Falls County: 20 feet, Limestone County: 40 feet, Navarro County: 1 foot, Williamson County: 50 feet.	8/11/2010
Queen City	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 12 feet; Fayette County Groundwater Conservation District: 60 feet; Lost Pines Groundwater Conservation District: 13 feet; Mid-East Texas Groundwater Conservation District: 0 feet; Post Oak Savannah Groundwater Conservation District: 30 feet.	8/11/2010
Simsboro (Middle Wilcox)	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 270 feet; Lost Pines Groundwater Conservation District: 237 feet; Mid-East Texas Groundwater Conservation District: 115 feet; Post Oak Savannah Groundwater Conservation District: 300 feet; Falls County: 0 feet, Limestone County: 43 feet, Navarro County: 1 foot, Williamson County: 55 feet.	8/11/2010
Sparta	Average drawdown between January 2000 and December 2059: Brazos Valley Groundwater Conservation District: 15 feet; Fayette County Groundwater Conservation District: 60 feet; Lost Pines Groundwater Conservation District: 7 feet; Mid-East Texas Groundwater Conservation District: 0 feet; Post Oak Savannah Groundwater Conservation District: 30 feet.	8/11/2010
Yegua-Jackson	Average drawdown from January 2010 to January 2060: Brazos Valley Groundwater Conservation District: 70 feet for the Yegua, 110 feet for the Jackson; Fayette County Groundwater Conservation District: 75 feet for the Yegua-Jackson; Post Oak Savannah Groundwater Conservation District: 100 feet for the Yegua-Jackson; Mid-East Texas Groundwater Conservation District: from January 2000 to January 2060, 5 feet for the Yegua-Jackson. Not relevant in Lost Pines Groundwater Conservation District.	6/30/2011
<b>GMA 13</b>		
Edwards (BFZ)	Maintain a minimum artesian flow of 500 gallons per minute from wells producing from the Edwards Aquifer in Frio County.	8/12/2010
Leona Gravel	Average drawdown of 15 feet in Medina County.	7/13/2011

Sparta, Weches, Queen City, Reklaw, and Carrizo-	Average drawdown of 23 feet.	4/9/2010
Yegua-Jackson	Average drawdown of 2 feet.	8/12/2010
<b>GMA 14</b>		
Brazos River Alluvium	Austin, Grimes, Waller, and Washington counties: from estimated 2010 conditions, the saturated thickness should be maintained at 90 percent. Not relevant in Brazos County.	8/25/2010
Carrizo Sand	Grimes County: from estimated 2010 conditions, the average drawdown should not exceed approximately 52.8 feet, Walker County: from estimated 2010 conditions, the average drawdown should not exceed approximately 45.7 feet.	8/25/2010
Gulf Coast	Listed DFCs by county and aquifer layer (Chicot, Evangeline, Burkeville, Jasper). See table below.	8/25/2010
Navasota River Alluvium	Grimes County: from estimated 2010 conditions, the saturated thickness should be maintained at 90 percent.	8/25/2010
Queen City	Grimes County: from estimated 2010 conditions, the average drawdown should not exceed approximately 16.8 feet, Walker County: from estimated 2010 conditions, the average drawdown should not exceed approximately 21 feet.	8/25/2010
San Bernard River Alluvium	Austin County: from estimated 2010 conditions, the saturated thickness should be maintained at 90 percent.	8/25/2010
San Jacinto River Alluvium	Walker County: from estimated 2010 conditions, the saturated thickness should be maintained at 90 percent.	8/25/2010
Sparta	Grimes County: from estimated 2010 conditions, the average drawdown should not exceed approximately 14 feet, Walker County: from estimated 2010 conditions, the average drawdown should not exceed approximately 19.5 feet.	8/25/2010
Trinity River Alluvium	Walker County: from estimated 2010 conditions, the saturated thickness should be maintained at 90 percent.	8/25/2010

Yegua-Jackson	Average drawdown from estimated 2010 conditions should not exceed, in Grimes and Walker counties: 10 feet in the unconfined Yegua, 15 feet in the confined Yegua, 20 feet in the brackish Yegua, 10 feet in the unconfined Jackson, 15 feet in the confined Jackson, 20 feet in the brackish Jackson, Polk County: 2 feet in the Yegua-Jackson, Washington County: 0 feet in the Yegua-Jackson. Not relevant in Jasper, Newton, and Tyler counties.	8/25/2010
<b>GMA 15</b>		
Gulf Coast	No more than 12 feet of average drawdown by 2060 relative to year 1999 conditions.	7/14/2010
<b>GMA 16</b>		
Gulf Coast	Average drawdown of 94 feet.	8/30/2010



## APPENDIX E – R CODE FOR CLEANING TEXT FILES AND FREQUENCY ANALYSIS AND FINDING CORRELATIONS

```
#### This code works for frequency and correlations
```

```
library(tm)
library(qdapDictionaries)
library(stringr)
library(SnowballC)
library(ggplot2)
library(qdap) # Quantitative discourse analysis of transcripts.
library(qdapDictionaries)
library(dplyr)
library(scales)
source("http://bioconductor.org/biocLite.R")
biocLite("Rgraphviz")
library(Rgraphviz) # Correlation plots.
library(SnowballC)
library(RColorBrewer)

install.packages("commonality.cloud")

#1961
corpus <- Corpus(DirSource("~/desktop/thesis/waterplans/text/." ), readerControl =
  list(language = "english"))

head(corpus)

corpus <- tm_map(corpus, content_transformer(stripWhitespace))
corpus <- tm_map(corpus, content_transformer(tolower))
corpus <- tm_map(corpus, removeWords, stopwords("english"))
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, PlainTextDocument)
corpus <- tm_map(corpus, removeNumbers)
corpus <- TermDocumentMatrix(corpus)
mcorpus <- as.matrix(corpus)
vecorpus <- sort(rowSums(mcorpus),decreasing=TRUE)
dcorpus <- data.frame(word = names(vecorpus),freq=vecorpus) ### Prepping the
dataframe of frequent words
inspect(corpus)

dfcorpus <- data.frame(head(dcorpus, 20))

ds <- DataframeSource(corpus)
corpus <- Corpus(VectorSource(corpus))

freqcorpus
```

```

N = 25
m <- as.matrix(corpus)
v <- sort(rowSums(m), decreasing=TRUE)
head(v, N) # Gives top N common words in my Corpus containing all state water plans

findAssocs(corpus, "water", corlimit = 0.95)

findAssocs(corpus, "plan", corlimit = 0.9)

findAssocs(corpus, "drought", corlimit = 0.978)

findAssocs(corpus, "planning", corlimit = 0.95)

findAssocs(corpus, "coal", corlimit = 0.95)

findAssocs(corpus, "debt", corlimit = 0.95)

d1961

corpus_clean <- tm_map(dtm1961, PlainTextDocument)

#####
toi <- "water" # term of interest
toi2 <- "drought"
corlimit <- 0.92 # lower correlation bound limit.
corlimitd <- 0.99
water_0.9 <- data.frame(corr = findAssocs(dtm1961, toi, corlimit)[[1]],
                        terms = names(findAssocs(dtm1961, toi, corlimit)[[1]]))
drought_0.9 <- data.frame(corr = findAssocs(dtm1961, toi2, corlimitd)[[1]],
                          terms = names(findAssocs(dtm1961, toi2, corlimitd)[[1]]))

water_0.9$corr <- factor(water_0.9$corr, levels =
  water_0.9$corr[order(water_0.9$corr)])

water_0.9$corr <- order(water_0.9$corr)
water_0.9$corr

require(ggplot2)
ggplot(water_0.9, aes( y = terms ) ) +
  geom_point(aes(x = corr), data = water_0.9) +
  xlab(paste0("Correlation with the term ", "\"", toi, "\""))

ggplot(drought_0.9, aes( y = terms ) ) +
  geom_point(aes(x = corr), data = drought_0.9) +
  xlab(paste0("Correlation with the term ", "\"", toi2, "\""))

#####
plot(d1961,
     terms=findFreqTerms(d1961, lowfreq = 100)[1:50],

```

```

cor = 0.5)

#####

#####

# 1968
wp1968 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/1968"), readerControl =
  list(language = "english"))
wp1968 <- tm_map(wp1968, stripWhitespace)
wp1968 <- tm_map(wp1968, content_transformer(tolower))
wp1968 <- tm_map(wp1968, removeWords, stopwords("english"))
wp1968 <- tm_map(wp1968, removePunctuation)
wp1968 <- tm_map(wp1968, PlainTextDocument) ## not including this in the script
dtm1968 <- TermDocumentMatrix(wp1968)
m1968 <- as.matrix(dtm1968)
v1968 <- sort(rowSums(m1968),decreasing=TRUE)
d1968 <- data.frame(word = names(v1968),freq=v1968) ### Prepping the dataframe of
  frequent words

df1968 <- data.frame(head(d1968, 120))
writeClipboard(df1968)

# 1984
wp1984 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/1984"), readerControl =
  list(language = "english"))
wp1984 <- tm_map(wp1984, stripWhitespace)
wp1984 <- tm_map(wp1984, content_transformer(tolower))
wp1984 <- tm_map(wp1984, removeWords, stopwords("english"))
wp1984 <- tm_map(wp1984, removePunctuation)
wp1984 <- tm_map(wp1984, PlainTextDocument)
dtm1984 <- TermDocumentMatrix(wp1984)
m1984 <- as.matrix(dtm1984)
v1984 <- sort(rowSums(m1984),decreasing=TRUE)
d1984 <- data.frame(word = names(v1984),freq=v1984) ### Prepping the dataframe of
  frequent words

df1984 <- data.frame(head(d1984, 120))

#1990
wp1990 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/1990"), readerControl =
  list(language = "english"))
wp1990 <- tm_map(wp1990, stripWhitespace)
wp1990 <- tm_map(wp1990, content_transformer(tolower))
wp1990 <- tm_map(wp1990, removeWords, stopwords("english"))
wp1990 <- tm_map(wp1990, removePunctuation)
wp1990 <- tm_map(wp1990, PlainTextDocument)
dtm1990 <- TermDocumentMatrix(wp1990)

```

```

m1990 <- as.matrix(dtm1990)
v1990 <- sort(rowSums(m1990),decreasing=TRUE)
d1990 <- data.frame(word = names(v1990),freq=v1990) ### Prepping the dataframe of
frequent words

df1990 <- data.frame(head(d1990, 120))
df

#1992
wp1992 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/1992"), readerControl =
list(language = "english"))
wp1992 <- tm_map(wp1992, stripWhitespace)
wp1992 <- tm_map(wp1992, content_transformer(tolower))
wp1992 <- tm_map(wp1992, removeWords, stopwords("english"))
wp1992 <- tm_map(wp1992, removePunctuation)
wp1992 <- tm_map(wp1992, PlainTextDocument)
dtm1992 <- TermDocumentMatrix(wp1992)
m1992 <- as.matrix(dtm1992)
v1992 <- sort(rowSums(m1992),decreasing=TRUE)
d1992 <- data.frame(word = names(v1992),freq=v1992) ### Prepping the dataframe of
frequent words

df1992 <- data.frame(head(d1992, 120))
df

#1997
wp1997 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/1997"), readerControl =
list(language = "english"))
wp1997 <- tm_map(wp1997, stripWhitespace)
wp1997 <- tm_map(wp1997, content_transformer(tolower))
wp1997 <- tm_map(wp1997, removeWords, stopwords("english"))
wp1997 <- tm_map(wp1997, removePunctuation)
wp1997 <- tm_map(wp1997, PlainTextDocument)
dtm1997 <- TermDocumentMatrix(wp1997)
m1997 <- as.matrix(dtm1997)
v1997 <- sort(rowSums(m1997),decreasing=TRUE)
d1997 <- data.frame(word = names(v1997),freq=v1997) ### Prepping the dataframe of
frequent words

df1997 <- data.frame(head(d1997, 120))
df

#2002
wp2002 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/2002"), readerControl =
list(language = "english"))
wp2002 <- tm_map(wp2002, stripWhitespace)
wp2002 <- tm_map(wp2002, content_transformer(tolower))
wp2002 <- tm_map(wp2002, removeWords, stopwords("english"))
wp2002 <- tm_map(wp2002, removePunctuation)
wp2002 <- tm_map(wp2002, PlainTextDocument)
dtm2002 <- TermDocumentMatrix(wp2002)
m2002 <- as.matrix(dtm2002)
v2002 <- sort(rowSums(m2002),decreasing=TRUE)

```

```

d2002 <- data.frame(word = names(v2002),freq=v2002) ### Prepping the dataframe of
frequent words

df2002 <- data.frame(head(d2002, 120))

#2007
wp2007 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/2007"), readerControl =
  list(language = "english"))
wp2007 <- tm_map(wp2007, stripWhitespace)
wp2007 <- tm_map(wp2007, content_transformer(tolower))
wp2007 <- tm_map(wp2007, removeWords, stopwords("english"))
wp2007 <- tm_map(wp2007, removePunctuation)
wp2007 <- tm_map(wp2007, PlainTextDocument)
dtm2007 <- TermDocumentMatrix(wp2007)
m2007 <- as.matrix(dtm2007)
v2007 <- sort(rowSums(m2007),decreasing=TRUE)
d2007 <- data.frame(word = names(v2007),freq=v2007) ### Prepping the dataframe of
frequent words

df2007 <- data.frame(head(d2007, 120))

#2012
wp2012 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/2012"), readerControl =
  list(language = "english"))
wp2012 <- tm_map(wp2012, stripWhitespace)
wp2012 <- tm_map(wp2012, content_transformer(tolower))
wp2012 <- tm_map(wp2012, removeWords, stopwords("english"))
wp2012 <- tm_map(wp2012, removePunctuation)
wp2012 <- tm_map(wp2012, PlainTextDocument)
dtm2012 <- TermDocumentMatrix(wp2012)
m2012 <- as.matrix(dtm2012)
v2012 <- sort(rowSums(m2012),decreasing=TRUE)
d2012 <- data.frame(word = names(v2012),freq=v2012) ### Prepping the dataframe of
frequent words

df2012 <- data.frame(head(d2012, 120))

#2017
wp2017 <- Corpus(DirSource("~/desktop/thesis/waterplans/text/2017"), readerControl =
  list(language = "english"))
wp2017 <- tm_map(wp2017, stripWhitespace)
wp2017 <- tm_map(wp2017, content_transformer(tolower))
wp2017 <- tm_map(wp2017, removeWords, stopwords("english"))
wp2017 <- tm_map(wp2017, removePunctuation)
wp2017 <- tm_map(wp2017, PlainTextDocument)
dtm2017 <- TermDocumentMatrix(wp2017)
m2017 <- as.matrix(dtm2017)
v2017 <- sort(rowSums(m2017),decreasing=TRUE)
d2017 <- data.frame(word = names(v2017),freq=v2017) ### Prepping the dataframe of
frequent words

```

```
df2017 <- data.frame(head(d2017, 120))
```

## **APPENDIX F – Leximancer Settings for Network Output**

- Theme size: 33%, Visible Concepts 0%, Gaussian and topical methods, combine word duplicates,
- General: Sentences per block: 2 (Normal)
- Prose Test Threshold: 0 (default)
- Duplicate Text Sensitivity: Auto
- Identify name like concepts: Yes
- Break at paragraph: Yes
- Auto-paragraphing: Yes
- Merge word variants: Yes

## REFERENCES

- Smith, A.E., Humphreys, M.S., 2006, Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior Research Methods*, 38 (2), 262-279
- Bastian, M., Heymann, S., 2009, Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*.
- Bender, J., Jimenez-Marroquin, M., Jadad, A. R., 2011, Seeking support on facebook: a content analysis of breast cancer groups, *Journal of Medical Internet Research*, v. 13.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003, Latent Dirichlet Allocation, *Journal of Machine Learning Research*, v. 3, p.993-1022.
- Blondel, V., et al., 2008, Fast Unfolding of Communities in Large Networks, *Journal of Statistical Methods*, v. 10, p. 1000.
- Brandes, U., 2001, A Faster Algorithm for Betweenness Centrality, *Journal of Mathematical Sociology*, v. 25, p. 163-177.
- Broek, P.V.D., 1995, A 'landscape' model of reading comprehension: Inferential Processes and the Construction of a Stable Memory Representation.: *Canadian Psychology/Psychologie canadienne*, v. 36, p. 53-54, doi: 10.1037/h0084723.
- Bruce, D. Newman, D., 1978, Interacting Plans. In *Cognitive Science*, Volume 2, Issue 3, July-September: 195-233
- Carley, K., 1993, Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis. In *Social Methodology*, Vol. 23: 75-126
- Cole-Lewis, H., Sanders, A., Varghese, A., Posada, S., Shwarz, M., Auguston, E., 2015, Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter, *Journal of Medical Internet Research*, doi: 10.2196/jmir.4969.
- Davis, A.P., 1917, *Irrigation Works Constructed by the United States Government*, New York: John Wiley & Sons Inc.
- Dovring, K., 1954, Quantitative Semantics in 18th Century Sweden: Public Opinion Quarterly, v. 18, p. 389, doi: 10.1086/266531.
- Dyer, G., 1983, The Role of Affect in Narratives. In *Cognitive Science* 7: 211-242,



Feinerer, 2008, An introduction to text mining in R., R News , 8(2):19., URL <http://CRAN.R-project.org/doc/Rnews/> .

Jacobi, C., Atteveldt, W.V., and Welbers, K., 2015, Quantitative analysis of large amounts of journalistic texts using topic modelling: Digital Journalism, v. 4, p. 89–106, doi: 10.1080/21670811.2015.1093271.

Lafferty, J., and Blei, D., 2009, Topic Models: Text Mining Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, p. 71–93, doi: 10.1201/9781420059458.ch4.

LERWPG, 2015, Llano Esacado (Region O) 2016 Regional Water Plan, Daniel B. Stephens & Associates, Inc.

Lind, S., 2011, Individual Level Studies – Micro, Sci2 Manual, <http://wiki.cns.iu.edu/display/SCI2TUTORIAL/5.1+Individual+Level+Studies+-+Micro>

Loyd, J., 2017, Monthly Rainfall (1968-2017), Online Publication, <http://etweather.tamu.edu/rainhistory/>

Mace, R. E., Petrossian, R., Bradley, R., Mullican, W. F., Christian, L., 2008, A Streetcar Named Future Desired Conditions: The New Groundwater Availability for Texas (Revised), Presented to the State Bar of Texas.

Majooni, A., Masood, M., and Akhavan, A., 2015, Scientific Visualizations Based on Integrated Model of Text and Picture Comprehension via Eye-tracking: Procedia - Social and Behavioral Sciences, v. 176, p. 52–59, doi: 10.1016/j.sbspro.2015.01.443.

McAllen, 2013, Community Environmental Scan, McAllen Chamber of Commerce, Report. <http://toolkit.mcallen.org/wp-content/uploads/2014/02/Environmental-Scan-2013.pdf>.

Nodus Labs. November 5, 2012, Texttexture: The Non-Linear Reading Machine. <http://noduslabs.com/cases/texttexture-non-linear-reading-machine/>

Ooms, Jeroen. 2016, Package ‘pdftools’. <https://cran.r-project.org/web/packages/pdftools/pdftools.pdf>

Pestian J., Nashrallah, H., Matykiewicz, P., Bennett, A., Leenaars, A., 2010, Suicide Note Classification Using Natural Language Processing: A Content Analysis, Biomedical Informatics Insights, v. 3, p. 19-28.

Pierce, S.A., Sharp, J.M., Guillaume, J.H.A. et al., 2013. Journal of Hydrogeology. 21: 331. doi:10.1007/s10040-012-0910-y

Rogers, E. W., Clancy, L., 2014, State Water Planning: Its History and its Challenges, Summer Conference – South Padre Island, June 20, 2014.

Tornberg, A., and Tornberg, P., 2016, Muslims in social media discourse: Combining topic modeling and critical discourse analysis: Discourse, Context & Media, v. 13, p. 132-142, doi: 10.1016/j.dcm.2016.04.003

TCEQ, 2014, Groundwater Regulation for Private Well Owners, [https://www.tceq.texas.gov/response/drought/groundwater\\_regulation.html](https://www.tceq.texas.gov/response/drought/groundwater_regulation.html).

TWDB, 2017, Groundwater Conservation Districts of Texas Map. [https://www.twdb.texas.gov/mapping/doc/maps/GCDs\\_8x11.pdf](https://www.twdb.texas.gov/mapping/doc/maps/GCDs_8x11.pdf).

TWDB, 2017, Groundwater Management Areas. [https://www.twdb.texas.gov/groundwater/management\\_areas/](https://www.twdb.texas.gov/groundwater/management_areas/).

TWDB, 2017, Major Aquifers. <http://www.twdb.texas.gov/groundwater/aquifer/major.asp>.

TWDB, 2017, State Water Planning. <https://www.twdb.texas.gov/waterplanning/swp/>.

U.S. Census Bureau, 2016, Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2016 (NST-EST2016-01), [www.census.gov/data/tables/2016/demo/popest/nation-total.html](http://www.census.gov/data/tables/2016/demo/popest/nation-total.html).

Waters, M., Wiersema, J., Stafford, T., 2008, A geoarchaeological evaluation of a reported late Pleistocene human burial from Brazoria County, Texas, Journal of Archaeological Science, v. 35, p. 2425-2433.

What is R?, 2017, The R foundation. <https://www.r-project.org/about.html>.

Wehner, J.E., 2011, Water Policy Informatics: A Topic and Time Series Analysis of the Texas State Water Plans [M.A. thesis]: The University of Texas at Austin. <https://repositories.lib.utexas.edu/bitstream/handle/2152/ETD-UT-2011-05-3537/WEHNER-THESIS.pdf>

Wise, J; Thomas, J; Pennock, K.; Lantrip, D.; Pottier, M.; Schur, A.; and Crow, V., 1995, Visualizing the non-visual: spatial analysis and interaction with information from text documents. From Information Visualization 1995 Proceedings: 51-58, 1995